Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : **2591-801X**

Content Available at : **http://journals.sfu.ca/jalt/index.php/jalt/index**

Critically reflecting on the use of immersive virtual reality in educational settings: What is known and what has yet to be shown?

Alyssa P. Lawson[A]

Amedee Marchand Martella[B]

[A] *University of California, Santa Barbara; Landmark College, Institute for Research and Training, USA*

[B] *University of California, Santa Barbara, USA*

Keywords

Immersive virtual reality;
learning environments;
learning technology;
virtual reality.

Correspondence

alyssaplawson@gmail.com [A]

Article Info

Received 12 September 2023
Received in revised form 18 October 2023
Accepted 23 October 2023
Available online 24 October 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.35

Abstract

Interest in the educational applications of immersive virtual reality (IVR) has continued to grow worldwide, particularly in recent years. With the ever-increasing literature base on IVR in educational contexts, two patterns of data have emerged: one focused on the affective component and one focused on the cognitive component of IVR. Research focused on the affective component of this technology has consistently found that it is beneficial in increasing students' motivation to learn. However, there is less of a consensus in the literature on the cognitive benefits of IVR, with results sometimes indicating it (a) is an effective tool for learning, (b) is *not* an effective tool for learning, and (c) is similar to other instructional media in its impact on learning outcomes. As suggested by these inconsistent findings, there is a great deal left to be understood about when and how IVR can be effective for learning. Therefore, the goal of this reflection article is to draw attention to important research gaps that, if filled, may help to explain the inconsistent effects of this technology in the research literature. Additionally, this article highlights areas in need of further research, which we hope will aid in the advancement of knowledge surrounding the effective implementation of IVR in education.

Introduction

To meet the demands of an ever-changing world, it is imperative that students develop critical academic competencies such as reading literacy, critical thinking, digital literacy, and math fluency. At the K-12 levels, the integration of technology in the classroom has been touted as a way to "accelerate, amplify, and expand the impact of effective teaching practices" (Office of Educational Technology, 2017, p. 5). At the college level, improving course instruction has been focused on moving away from the traditional lecture method and towards methods that provide students with opportunities to be more involved in the learning process—termed *active learning* (Association of American Universities, 2017). Immersive virtual reality (IVR) is one tool that has been a topic of interest in the fields of education, educational psychology, and educational technology given that it moves instruction away from more passive forms of learning. IVR does this by fully immersing learners in a new environment, often through a head-mounted display that presents screens in front of each eye. Typical IVR devices include Oculus Quest, PlayStation VR, HTC Vive, and Google Cardboard, but there are many other devices currently available or in development. The immersion afforded by IVR devices allows students to feel as though they are in an environment different from the one in which they are physically present. By immersing students in interactive learning environments, they often increase their sense of presence and agency; these two factors can lead to increases in interest, intrinsic motivation, self-efficacy, embodiment, and self-regulation, which in turn are useful in improving their learning outcomes (see the *Cognitive Affective Model of Immersive Learning* by Makransky & Petersen, 2021).

Despite the increasing popularity of IVR and the appeal of using a technology that, in theory, can benefit students both cognitively and affectively, the results of empirical investigations comparing IVR to other media or to more traditional forms of instruction are mixed, leading to a number of questions as to when and how IVR can be effective for student learning. These mixed results in the literature make it difficult to provide specific recommendations to educators on when and how to effectively implement IVR in the classroom and to provide specific guidelines to VR designers on how to create effective learning environments within an IVR experience.

Therefore, the purpose of this reflection article is to discuss several research gaps that, if filled, may help to explain the inconsistent effects of IVR and to provide subsequent research directions aimed at advancing knowledge surrounding the effective implementation of IVR in education. Before discussing what more needs to be understood about IVR technology, we will first provide a brief overview of research on IVR.

Research on immersive virtual reality

Immersion is one of the prominent aspects of IVR that distinguishes it from desktop virtual reality (DVR; an interactive virtual world presented on a desktop screen) and other types of traditional instructional methods such as lectures/educational videos (Makransky, 2020; Makransky & Petersen, 2021). Immersion can be considered the extent to which the system creates a new virtual world for the learner. When immersion increases—and thus the virtual world is more vivid and realistic—a learner experiences higher presence, which is the subjective experience of being in the environment. Presence is important to develop because an increase in presence can increase certain affective characteristics, like interest and motivation (Makransky, 2020; Makransky & Petersen, 2021). By making students feel as though they are really in the new environment, they likely will enjoy the lesson more, which can increase their motivation to learn and their focus on the task. This may ultimately lead to a deeper understanding of the material.

The motivational and affective benefits of IVR lessons have been well researched and established within the literature. Indeed, empirical research studies and reviews of these studies often demonstrate that IVR is beneficial to the affective experience of learners, with positive effects shown across studies that vary in implementation style, the topic being taught, and the research design used. For example, IVR can positively impact motivation (e.g., Akgün & Atici, 2022; Matovu et al., 2023; Parong & Mayer, 2018 [Experiment 1]; Villena-Taranilla et al., 2022; Yu, 2021), interest (e.g., Akgün & Atici, 2022; Flavia Di Natale et al., 2020; Makransky et al., 2020; Parong & Mayer, 2018 [Experiment 1]; Yu, 2021), and self-efficacy (e.g., Akgün & Atici, 2022; Makransky et al., 2020), with benefits consistently shown across studies that vary in implementation style, the topic being taught, and the experimental design used.

Although the literature on motivation and IVR lessons has been consistent across studies, the research on the benefits of IVR lessons on *learning outcomes* has not demonstrated clear results (Matovu et al., 2023). On one side of the spectrum, many research studies have shown benefits of presenting learning material in an IVR lesson on certain learning outcomes (e.g., Alhalabi, 2016; Kim et al., 2019 [for long-term memory scores]; Kozhevnikov et al., 2013; Makransky et al., 2019a [for behavioral transfer tests]; Webster, 2016; Yang et al., 2022). For example, Makransky et al. (2019a) tasked students with learning lab safety by text, DVR, or IVR. This study demonstrated that students did better on behavioral transfer tests (i.e., lab safety tests conducted in the real world) when they learned in IVR compared to when they learned via text. There have also been a handful of meta-analyses published recently that have demonstrated a positive, albeit small, impact of learning from IVR. For example, in Coban's (2022) review of 49 primary studies on IVR, the overall effect size of using IVR in learning was positive but small (g = .38; Hedge's g is a common metric used to measure the magnitude of difference between two groups), with the effect being strongest in fields like architecture, engineering, geometry, and chemistry. Similarly, Wu et al. (2020) analyzed 35 studies and found that the overall effectiveness of IVR was also positive but small (g = .24).

On the other side of the spectrum, some researchers have found that IVR leads to significantly lower performance on certain learning outcomes as compared to other instructional media/approaches (e.g., Makransky et al., 2019b [for knowledge test]; Makransky et al., 2021 [Experiment 2, for declarative knowledge test]; Moreno & Mayer, 2004 [for retention test]; Parong & Mayer, 2018 [Experiment 1, for factual questions], 2021a [for transfer test], 2021b [for transfer test]). When compared to more traditional or 2D formats, IVR has proved less useful for students' learning in several studies. For example, in multiple studies in which students learned about cells and the blood stream, students who learned using IVR performed significantly lower on factual questions (Parong & Mayer, 2018 [Experiment 1]) and transfer test (Parong & Mayer, 2021a) than those who learned using a slideshow. When compared to non-immersive technologies, IVR also has proved less useful for students' learning; for example, in a game lesson about designing plants for different environments, participants who learned with IVR did not perform as well as those who learned from DVR on both the retention test (Moreno & Mayer, 2004).

In the middle of the spectrum, a large portion of research studies have found no significant difference in the effectiveness of IVR on certain learning outcomes when comparing IVR to other types of instructional media (Ekstrand et al., 2018; Hassenfeldt et al., 2020; Liu et al. 2021; Makransky et al., 2019a [for retention test]; Makransky et al., 2019b [for transfer test]; Makransky et al., 2021 [Experiment 1; Experiment 2, for procedural knowledge and transfer tests]; Parong & Mayer, 2021a [for retention test], 2021b [for retention test]. When looking to systematic reviews for a more comprehensive view of the literature, Luo et al., (2021) meta-analyzed 22 articles from 2000 to 2019 with HMD as a moderator and found that the overall effect of using IVR was not meaningfully different than other instructional media (g = .20, with the 95% confidence interval ranging from an effect size of -.16 to .55).

Given these inconsistent results found within the IVR literature, there is a great deal left to be understood about IVR technology surrounding when and how it can be effectively implemented to promote student learning. Therefore, it is important to discuss research gaps that, if filled, could help to explain the varied results, may help to establish boundary conditions for when IVR is effective versus when more traditional forms of instruction are most useful to students, and could provide insight into how to design effective IVR lessons. Gaps in the literature and subsequent research directions aimed at filling these gaps will be discussed in more detail in the following section.

What more needs to be understood about IVR technology?

To better understand when and how IVR technology is effective for student learning, we (a) consider several possible research gaps that, if filled, may help to explain the inconsistent learning results within the IVR literature and (b) present subsequent research directions that we believe are important for advancing knowledge on how to effectively implement IVR in education. We will focus on three research gaps that we deem to be most pressing, although there are other gaps that exist within the literature. The first research gap relates to whether IVR imposes larger demands on working memory than other forms of instruction. The second research gap relates to how IVR lessons are being designed in experimental studies. And the third research gap relates to what IVR is being compared to in experimental studies. Each research gap is discussed below.

Research gap 1: The cognitive load of the IVR lesson

The impact of the cognitive load of an IVR lesson reflects a research gap within the IVR literature. There are two parts of this research gap that will be discussed: the demands that are imposed on working memory during a lesson and the type of cognitive load that is being increased during a lesson.

Research gap 1A: Demands on working memory

The cognitive load of a lesson can vary from one instructional approach to another. Cognitive load refers to the demands imposed on working memory during learning. Working memory is severely limited in the amount of information that can be processed at one time, which is essential to understand in terms of its role in learning and instruction (Fenesi et al., 2015). With working memory only able to process so much novel information at one time, it is vital for students to be able to deal with all of the novel incoming information. Unfortunately, how IVR lessons impact learners' cognitive load, as compared to other instructional methods, remains an under-researched area.

For those studies that have examined how IVR lessons impact learners' cognitive load (e.g., Baceviciute et al., 2021; Huang et al., 2021; Makransky et al., 2019b, Mayer et al., 2022; Parong & Mayer, 2018, 2021a, 2021b; Petersen et al., 2022), results tend to show that there is an increase in cognitive load when learning from IVR. For example, Makranksy et al. (2019b) had students learn how to conduct a lab procedure in a chemistry course using either IVR or a 2D computer simulation. These learning mediums were then switched in the second part of the lesson. During these lessons, students' cognitive load was measured using EEG. During the first part of the lesson, there were no differences in the amount of workload participants experienced but in the second part of the lesson, those in the IVR condition had higher load than those on the computer. These results demonstrate that IVR can add to students' cognitive load, particularly if they are asked to learn additional material in IVR after already viewing another lesson. Similarly, in a lesson on cells and the bloodstream where students learned via IVR or a slideshow, participants in the IVR condition reported that they were more distracted and/or had a harder time focusing during the lesson as compared to those who viewed the slideshow (Parong & Mayer, 2018).

Not all IVR lessons are designed in the same way, which may lead to variations in the amount of cognitive load that is imposed on learners. Subsequently, certain IVR lessons could impose larger demands on working memory than the instructional methods to which they are being compared. Similarly, because not all instructional methods to which IVR lessons are compared are designed or implemented in the same way, they may also vary in the amount of cognitive load imposed on learners. These possibilities could lead to inconsistent findings in the IVR literature depending on how each mode of instruction has been designed. Being more mindful of the design of the lessons, both within IVR and outside of IVR, and examining the demands that are being placed on working memory during learning would help to identify any potential barriers for learners' processing of relevant information and lend insight into why there might have been differential effects of the compared approaches.

Future research should further investigate the impact of the cognitive load caused by IVR. Cognitive Load Theory (CLT; Paas & Sweller, 2022; Sweller, 1994, 2020) and Cognitive Theory of Multimedia Learning (CTML; Mayer, 2022) are useful theories to consider when specifically designing instruction in IVR. With more traditional forms of instruction (e.g., watching an instructor present information or watching an instructional video on a computer screen), all of the relevant information students need to focus on is presented directly in front of them in one place. With this directly presented information, students can focus their attention and more easily recognize information that is distracting and irrelevant to the lesson. However, in IVR lessons, information that students may be asked to focus on comes from all directions, potentially putting students in a situation where they could easily miss key ideas. Therefore, future research should specifically focus on how the design of the different modes of instruction being compared in a study affect students' cognitive load and learning outcomes. Further, researchers should examine how individual differences in learners' cognitive skills could impact how well they learn with different types of media (Lawson & Mayer, under review) and how these differences could impact the effectiveness of IVR. Given that learning new material through IVR can tax students' limited working memory via the presentation of extraneous material, it is important to examine how this technology affects students who may be able to handle this additional cognitive load (through better inhibition ability or being able to ignore distractions) versus students who are less able to handle this cognitive load (see Albert et al., 2020 and Grenell & Carlson, 2021 for a discussion of individual differences in executive function and academic achievement/learning).

Research gap 1B: Type of cognitive load being increased

Another glaring issue related to cognitive load is the minimal investigation by researchers into what *type* of cognitive load is being increased during a lesson. Researchers have investigated and differentiated three different types of cognitive load that stem from instructional material (Mayer, 2022; Paas & Sweller, 2022; Sweller, 1994, 2020). Extraneous load or extraneous processing occurs when information is presented in a lesson that is not relevant to the lesson itself, such as irrelevant facts or pictures that draw learners' attention away from the important information being conveyed. Intrinsic load or essential processing occurs when learners build a mental representation of the presented material; this load increases as the complexity or the interactivity of the material increases. Lastly, some researchers propose a third kind of processing called germane load or generative processing that occurs when a learner works to make sense of the material presented, develops connections between different parts of the material, and connects the novel information with their prior knowledge.

These different types of cognitive load vary in how they impact learning—having an increase in extraneous processing would likely hurt students' learning because learners are paying attention to information that is not relevant to the main goal of the lesson whereas an increase in generative processing would likely benefit students' learning because it encourages deeper processing of the learning material. For example, in one study conducted by Parong and Mayer (2021a), participants' different levels of cognitive load were measured using self-report Likert scale questions and their reported cognitive state was measured through EEG. Participants reported having higher cognitive load, *specifically extraneous processing*, when they learned in the IVR condition compared to when they learned in the slideshow condition. Furthermore, a mediation analysis of the relationship between learning condition, *extraneous processing*, and retention scores found that there was a significant mediation path wherein those in the IVR condition reported higher extraneous processing which in turn led to worse retention scores as compared to those in the slideshow condition.

Without differentiating the type of processing that a learner is experiencing, as is often an issue in the literature, it is difficult to determine whether increases or decreases in cognitive load will be helpful or detrimental to student learning. Therefore, the question of what type of cognitive load drives the increase in reported cognitive load from IVR technology needs to be further investigated to better to determine why certain patterns of data are emerging from IVR studies.

Future research should be focused on how different components of learning in IVR contribute to different types of cognitive load and what types of strategies can reduce less desirable types of cognitive load (like extraneous processing) and increase more desirable types of load (like generative processing). As part of this research direction, we need to understand what components of an IVR lesson (e.g., interacting with objects in the virtual environment, experiencing a large amount of visual information, trying to understand how to use the device) impact different types of cognitive load. With this understanding in place, we can then better understand how to help remedy the issues that cause increases in extraneous processing and focus our attention on how to encourage generative processing during a lesson.

Research gap 2: The impact of how IVR conditions are designed in IVR studies

The impact of how IVR conditions are designed in IVR studies reflects a research gap within the IVR literature. There are two parts of this research gap that will be discussed: the prioritization of learning theory in the design of an IVR lesson and whether IVR is used to exclusively teach content or is used as a supplemental instructional tool.

Research gap 2A: Learning theory prioritization

Just as much of the literature on learning with IVR lessons does not often consider the demands imposed on working memory during learning, nor does it always prioritize learning theories in the design and implementation of this technology in an educational setting (Lui et al., 2023; Matovu et al., 2023; Radianti et al., 2020). Indeed, a gap in the literature exists for whether the alignment of IVR lessons to effective design guidelines differentially impacts the technology's effectiveness. If certain IVR lessons follow better design guidelines than other IVR lessons, it is highly possible that the outcomes of using these various lessons for learning could impact results, particularly if these lessons lead to differences in how students cognitively interact with the material presented in the lesson. These design considerations could impact the various types of cognitive load present in the IVR lesson, as previously discussed, with learners potentially struggling to keep up with the information presented, thereby hindering learning. These design considerations could also impact the degree of cognitive engagement being cultivated during a lesson.

There have been some studies that have involved the incorporation of learning theory into the design of IVR lessons. More specifically, several researchers have integrated generative learning strategies into IVR to reduce extraneous processing and manage essential processing and/or increase generative processing (e.g., Klingenberg et al., 2020; Makransky et al., 2021; Parong & Mayer, 2018 [Experiment 2]). However, this research area is small and is in need of further investigation. Therefore, there is a need for more focused research on the incorporation of learning theory into the design of IVR lessons to continue as this can provide more insight into how to induce learning more effectively through IVR lessons. It may be the case where IVR lessons that are designed based on learning theories and include effective learning strategies are more effective for student learning than those that do not involve these design considerations, which may be contributing to the variability in findings across the IVR literature.

In future research, investigators should apply learning theories to IVR lessons in the pursuit of recognizing what aspects of the IVR lesson promote better learning outcomes. They should also investigate whether certain learning strategies are more effective during an IVR lesson than other strategies. For example, perhaps adding self-explanations to an IVR lesson helps students learn more than adding retrieval practice activities to the lesson. Or perhaps these strategies work best in tandem—that is, both are needed to maximize student learning. Further, when incorporating effective learning strategies into IVR lessons, researchers should assess whether strategies that have been demonstrated to be effective for learning, such as practice tests, peer teaching, self-explanations, feedback, supplemental instruction, etc., confer the same benefits when embedded in traditional forms of instruction versus when embedded in IVR lessons. In other words, is it the VR technology and immersion themselves that are key in promoting learning, or is it simply the fact that students benefit from embedded strategies that have been demonstrated to be effective for learning, regardless of the particular mode of instruction used during a lesson? It could be that traditional forms of instruction are just as effective as IVR when combined with effective learning strategies. In other words, strong IVR conditions should be compared to strong comparison conditions to determine if IVR provides learning benefits above and beyond the learning strategies themselves (a point elaborated on under Research Gap 3).

Research gap 2B: Exclusive IVR or integrated IVR?

When examining the IVR conditions in IVR studies, one will find that some studies use IVR to directly teach students content (e.g., Lui et al., 2020; Madden et al., 2020; Parong & Mayer, 2018, 2021a, 2021b; Su et al., 2022) whereas other studies use it as an extension activity that is integrated into more traditional modes of instruction (e.g., Campos et al., 2022; King et al., 2022; Liu et al., 2022; Stranger-Johannessen, 2018). These two uses of IVR highlights an important question: Is IVR more effective when used to directly teach content or when used as an active learning activity that is embedded in more traditional forms of instruction? IVR may be most useful as an extension activity to help students further encode the content and promote generalization programming through immersive activities. For example, Lui et al. (2020) taught students about the *lac* operon in two 80-minute lectures. One group of participants received IVR lessons over the following two weeks to help reinforce the ideas from the lecture while the other group did not. Students who participated in the IVR session learned more about the *Lac* Operon Concept Inventory than those who did not participate in this session.

As another example, Makransky & Mayer (2022) studied the impact of a six-lesson intervention that involved teaching students about climate change. The first lesson for both the IVR condition and the video condition involved a fake news article followed by a discussion of controversy surrounding climate change. The second lesson involved instruction on the scientific method and a virtual field trip to Greenland. During this lesson, students in the IVR condition experienced a 360-degree virtual trip to Greenland whereas students in the comparison condition experienced the virtual field trip to Greenland as a class video that was displayed on a projector screen. After this lesson, they took an immediate posttest. Subsequent lessons for both conditions involved learning about experiments and interpreting results as regular class sessions, and a delayed posttest was given after the last session. Results indicated that students learned more when they were able to take the virtual field trip in conjunction with the other course material presented in regular class sessions as compared to those who watched the video in conjunction with the other course material presented in regular class sessions, both on the immediate and delayed posttest. Therefore, using IVR within a more traditional class context seemed to improve student learning.

Within the active learning literature, similar benefits have been demonstrated for the integration of active learning activities into traditional STEM classes (see Freeman et al., 2014). Perhaps when used as a motivational tool and/or an additional encoding tool paired with class lectures or other instructional videos/materials, IVR more consistently improves learning than when it is used to exclusively teach class content. This idea of using IVR as a supplemental tool in education needs to be directly tested, particularly against exclusive VR conditions, before making conclusions about its benefits as a learning tool.

In future research, it is important to assess whether IVR, which is used to directly teach content, is more or less effective than when it is used as an active learning activity embedded in more traditional forms of instruction. Researchers should examine the effects of students learning content directly through an IVR lesson or through an integration of IVR technology and more traditional modes of instruction such as class lectures. Given the potential issues with distracting information in IVR lessons and the subsequent increases in extraneous processing, providing students with foundational knowledge outside of IVR may help to reduce the demands imposed on working memory. As an example, when researching the integration of IVR and other modes of instruction, researchers should also investigate how much class time should be dedicated to IVR experiences and whether this tool should occur after a lecture or should be interspersed throughout the lecture (see Martella & Schneider, in press, for information on lecture and active learning integration). These types of considerations would provide direction to instructors looking to implement the technology in their classroom.

Research gap 3: The impact of how comparison conditions are designed in IVR studies

The impact of how comparison conditions are designed in IVR studies reflects a research gap within the IVR literature. Two parts of this research gap will be discussed: (a) the type and quality of the comparison conditions and (b) the extent to which causal conclusions can be made regarding the efficacy of specific component(s) of the treatment package.

Research gap 3A: Type and quality of non-VR comparison conditions

The impact of how comparison conditions are designed in IVR studies reflects a research gap within the IVR literature. Indeed, the impact of the type and quality of these conditions has not been well studied. One potential issue when designing intervention studies, particularly in the context of instructional comparisons, is the inclusion of “strawman” conditions—conditions that are “easy to knock down” or, in other words, are doomed to fail from the start. For example, inquiry-based conditions that are unguided can be described as a “strawman” in that they are unlikely to be effective and do not serve as a fair comparison to alternative methods, such as direct instruction, given what we know about the importance of guided instruction (see Davis et al., 2017; Hmelo-Silver et al., 2007). Traditional lecture conditions can also serve as a "strawman” if the lecture is fully passive, poorly presented, and the slides are convoluted and/or involve extraneous information. To make the lecture a better comparison condition, strategies such as using mental imagery or increasing the structure of a lecture through outlines, for example, can be incorporated during the lecture design phase (deWinstanley & Bjork, 2002) as can taking into account multimedia design principles such as not presenting the same information in multiple formats simultaneously (Mayer, 2022).

Unfortunately, the extent to which comparison conditions reflect fair comparisons in IVR studies has not been well examined by researchers nor has the impact of different forms of “traditional” instruction. When reading the literature (both published and unpublished studies), the variation in what IVR lessons are being compared to is vast, with conditions involving 2D static images (Porter et al., 2019), class lectures/PowerPoint lessons/recorded videos (Bukoski, 2019; Drake, 2022; Lamb et al., 2018; Parong & Mayer, 2018; Sanzana et al., 2022), textbook/booklet/manual readings (Alrehaili & Al Osman, 2022; Makransky et al., 2019a; Tarng et al., 2022), and hands-on activities (Greenwald et al., 2018; Madden et al., 2020), among a myriad of other modes of instruction. This variation is, unfortunately, collapsed across conditions in meta-analyses, perhaps clouding the impact of different types of comparison conditions on student learning as compared to IVR conditions.

A closer reading of these conditions illuminates the presence of many “strawmen” that are unlikely to be effective as compared to the IVR conditions. For example, in one study, participants in the control group received up to 45 minutes to study a 14-page lab manual on lab safety that was designed according to instructional design principles; participants in the IVR condition were engaged in many activities involving narrative guidance, feedback, practice multiple-choice questions, and lab safety tasks to perform (Makransky et al., 2019a). The retention assessment for both conditions involved multiple-choice questions testing for conceptual and procedural knowledge. The transfer assessment included behavioral tasks—testing experiences that drew on experiences practiced in the IVR condition but not in the control condition. The IVR condition and text condition performed similarly on the retention test but unsurprisingly, the IVR condition resulted in significantly higher scores on the behavioral transfer tasks than the control condition. But one must ask whether the comparison condition served as a fair control. It could be argued that the increase in learning in IVR was not due to the immersive nature of the lesson but instead due to (a) a testing/practice effect, (b) the embedded strategies that the control condition did not receive, (c) more structured learning, and/or (d) the fact that students in the control condition were not explicitly taught the content.

As another illustrative example, learners in a study conducted by Tarng et al. (2022) were asked to learn about physics concepts related to time and space travel. The IVR condition learned this content by playing a pilot and control room staff to complete space exploration missions. During this lesson, they were asked questions that they needed to answer correctly to move forward in the lesson and were also given real-time feedback. The comparison group learned the content by reading a physics textbook. Although the study concludes that the IVR lesson was more effective than the textbook lesson, it is important to discuss whether the driving force for these learning differences was the active learning experiences and practice questions learners received during the lesson or whether it was the immersive nature of the lesson.

Unfortunately, the nature of the conditions compared in IVR studies has not been subjected to critical analysis. Ideally, to determine whether schools should spend money on VR equipment and whether instructors should take the time to adopt and implement the technology in their classrooms, the comparison conditions that serve as control conditions should reflect true “business-as-usual” or “regular instruction” conditions in that what students are asked to do in these studies reflects what they would/could really do during classroom learning. Though some may argue that students are typically asked to read textbooks for class, for example, it is important to ask whether the only exposure students receive to the content in class is through textbook readings. More often, the readings are assigned to provide exposure before coming to class or to solidify learning after class and are not used in isolation, as has been found in IVR studies.

IVR lessons should also be compared to well-designed alternative approaches such as interactive lectures or validated curricular materials that involve effective learning strategies. Designing passive control conditions such as traditional lectures is no longer productive (Freeman et al., 2014). As touted in the active learning literature, incorporating more opportunities for students to participate in the learning process can be beneficial to their learning (see Freeman et al., 2014). Although the active learning literature suffers from many of the same issues as the IVR literature, there is a great deal of research on the benefits of engaging students in learning activities such as retrieval practice, elaborative interrogation, and self-explanations (Dunlosky et al., 2013). Perhaps students do not need to be immersed in a virtual world to experience boosts in motivation and learning—it may simply be the case that we need to design more interactive learning experiences within the real-world class context to aid students in their learning.

By continuously comparing IVR conditions to bad control conditions, we overlook the potential benefits of this technology for real-world instruction and miss out on the opportunity to offer specific suggestions to instructors on how to improve classroom instruction. As an example of a fair comparison condition, Petersen et al. (2022 [Experiment 2]) compared an IVR lesson with an active pedagogical agent who taught a lesson about pipetting in a virtual laboratory setting to a real-life lesson with an instructor who taught a lesson about pipetting in a chemistry laboratory. The setup between conditions was designed to be identical and students in both conditions received active practice with pipetting in addition to explicit instruction. Results indicated that students in the IVR condition made more errors in dexterity with the pipette but performed similarly for serial dilution and safety performance on the real-life transfer test, had lower declarative knowledge scores, experienced significantly higher extraneous cognitive load, and had smaller increases in self-efficacy than the real-life condition. Although the discussion is largely framed to justify the use of VR as a complement to traditional teaching, it is important to highlight that traditional teaching served as an overall better intervention than IVR when designed to be comparable in terms of effective learning experiences (e.g., explicit instruction and active practice).

One critical direction for researchers to take in future research studies is to determine the extent to which comparison conditions reflect fair comparisons in IVR studies, perhaps through a systematic review of studies that have already been conducted. Further, researchers should design more studies that involve true “business-as-usual” conditions or that include rigorous instructional practices (e.g., interactive lectures) and materials as comparison conditions, such as the example above by Petersen et al. (2022 [Experiment 2]). Finally, researchers should investigate the impact of different forms of instruction to determine if specific comparison conditions promote greater student learning than other types of comparison conditions. It may be the case where well-designed lectures serve as a better learning medium than well-designed videos, for example.

Research gap 3B: Causal conclusions about why one condition was more effective

Another important consideration when designing comparison conditions in IVR studies is whether these conditions allow researchers to make causal conclusions about *why*, specifically, one condition outperformed (or did not outperform) the other. To be able to make valid inferences, an experiment should be unconfounded, with only a single contrast occurring between conditions (Klahr, 2013). IVR conditions are often designed as a treatment package with many different instructional components and are frequently compared to a control condition that gets a different, albeit minimal, treatment package (e.g., Alrehaili & Al Osman, 2022; Chittaro & Buttussi, 2015; Makransky et al., 2019a; Tarng et al., 2022). We adopt the use of “treatment package” and “components” from Ward-Horner and Sturmey (2010) in that we use “*component* to refer to variables that comprise a *treatment package* and treatment package to refer to the application of an intervention with all of its components” (p. 686). At first glance, an IVR treatment package versus a control treatment package reflects one contrast—the type of treatment package students receive. However, upon closer examination, the instructional components of these treatment packages are generally different between conditions, leading to more than one contrast between conditions.

As an illustrative example, Alrehaili & Al Osman (2022) assigned students to an IVR condition, a DVR condition, or a booklet condition to assess the impact of immersion. The IVR condition was designed according to multimedia principles. It involved a tutorial video to introduce students to relevant concepts and four different game levels with contextual guidelines in the form of textual messages and with specific tasks to complete. Each level of the game built upon the previous level. The comparison conditions received either the IVR lesson given via a computer monitor rather than a head-mounted display or a small booklet on honeybees that was written to mimic a 7th or 8th grade textbook. This booklet included many pictures from the IVR honeybee game. These conditions reflect three different treatment packages, each with different instructional components such as tasks, videos, and textual messages in IVR and written facts and pictures in the booklet, making it difficult to interpret outcomes on the knowledge test.

When there are multiple differences between or among conditions, only general conclusions can be drawn; for example, "IVR was more effective than the traditional condition." This conclusion might be satisfactory to some researchers who want to know if IVR is more effective than business-as-usual instruction, for example. However, *why* the IVR condition was more effective than the traditional condition cannot be answered from studies that involve multiple differences. It may be the case where immersing students in a virtual environment is the causal factor, or it may be the case where the other instructional components that were embedded in the IVR lesson but not in the comparison condition were responsible for boosting student performance.

Therefore, before we can conclude that immersing students in a virtual world is necessary for improving student learning and motivation beyond what we can give them in a traditional classroom, we need to be sure it is, in fact, the immersion and/or interactivity afforded by the VR environment and not the added instructional components embedded in an IVR lesson that result in greater learning gains than comparison conditions. Without knowing why an IVR condition resulted in greater learning, it is difficult to offer specific guidance and practical advice on effective IVR implementation to instructors (see similar discussion in Martella & Schneider, in press) and to definitively say that IVR should be adopted by instructors as compared to more affordable, real-world instructional interventions. The comparison conditions should thus be designed intentionally to minimize differences between the conditions being compared in the study.

In continuing to do research on IVR, it is vital to isolate and compare key instructional features. There have been a number of research studies that have isolated the immersion and interactivity component of IVR conditions by comparing an IVR lesson to a less immersive DVR lesson (e.g., Alrehaili & Al Osman, 2022; Barnidge et al., 2022; Makransky et al., 2019a) or a more passive 2D video lesson (e.g., Allcoat & von Mühlenen, 2018; Parong & Mayer, 2021a). Although these studies do control variables and isolate the impacts of the IVR technology, they typically afford little insight into how classroom instruction compares to IVR lessons when variables are controlled and contrasts are kept to a minimum (see Petersen et al., 2022 [Experiment 2] for an example of how classroom instruction can be compared to IVR while minimizing differences). Therefore, future research should expand on these prior studies by using component analysis to compare IVR and non-VR conditions, particularly those that incorporate effective learning strategies so as not to fall into the "strawman" trap. One way to identify active elements is to conduct a factorial design where two variables are examined via four conditions, for example. As an illustrative example, Parong and Mayer (2021a) examined whether an IVR lesson was as effective as an equivalent PowerPoint lesson on a desktop computer and whether the generative learning strategy of practice testing boosted performance in either medium. Their study involved a 2 X 2 factorial design with four conditions: IVR, PowerPoint, IVR + practice testing, and PowerPoint + practice testing. Results indicated students who received the IVR lesson performed significantly lower on the transfer test and performed lower, albeit not statistically significantly lower, on the retention test than those who received the PowerPoint lesson, with or without practice questions added to the lessons. This study, therefore, lends insight into the impact of the instructional medium as well as the impact of an embedded generative learning strategy.

Discussion

Recommendations for the future of IVR

In this reflection, we have highlighted three main areas in need of further investigation to advance knowledge surrounding the effective implementation of IVR in education. These include (a) a more thorough investigation of the impact of IVR lessons on different types of cognitive load, (b) a deeper look into how IVR conditions are being designed and whether design differences impact results, and (c) a deeper look into how comparison conditions are being designed and whether design differences impact results. Based on these research gaps, we have put together a set of recommendations for researchers, educators, and VR developers that reflect what we currently know about the field as well as where we believe future research should go next.

Learning and technology researchers

Researchers have the fundamental job of bridging the gap between the development of VR technology and the effective implementation of this technology in the classroom. Although there is a growing literature base on IVR in educational contexts, this reflection demonstrates that there are many gaps remaining that need to be filled in order to determine the usability of this technology for education. As discussed throughout our reflection, there are three major gaps in the literature that need to be addressed in order to determine whether and how best to use this technology in education.

Regarding cognitive load, there are two primary areas in which researchers in technology and learning should further investigate. First, we recommend that researchers investigate how different types of IVR lessons impact learners' cognitive load while learning. For example, researchers could investigate how changes in the interactivity of an IVR lesson can impact the cognitive load learners experience during the lesson. Second, we recommend that researchers keep in mind the variation in types of cognitive load and subsequent impacts on learning when conducting research on the impact of cognitive load in learning with IVR technology. For example, researchers should investigate how learning in a specific IVR lesson can increase or reduce learners' extraneous, essential, and generative processing.

Regarding the integration of IVR into classroom environments, there are two primary areas that should be further investigated. First, we recommend that researchers incorporate learning theories into the designs of lessons being used in research that investigates the use of IVR technology in learning. For example, researchers could design IVR lessons according to Bloom's Taxonomy (Bloom, 1956; Anderson & Krathwohl, 2001) to target different types of knowledge during the lesson. Second, we recommend that researchers investigate the benefits of using IVR as a way for exclusively learning content (i.e., learning content with IVR lessons only) compared to as a way to support learning (i.e., as an active learning tool embedded in a more traditional classroom structure). For example, researchers could directly compare how well students learn content when they are taught the content directly with IVR (including explicit instruction and practice) versus when they are taught the content outside of IVR (e.g., in a more traditional type of learning environment) and use IVR to help cement the new skills through immersive practice activities.

Regarding the comparison conditions used in IVR research, there are two primary areas that are in need of deeper investigation. First, we recommend that researchers investigate the impact of using strawmen conditions versus more well-developed control conditions in IVR research. For example, researchers could investigate the impacts of using IVR for learning when compared to a condition in which students simply read a textbook or when the activities in the control condition match those that are done in the IVR lesson. Second, we recommend that researchers think critically about the research question under investigation when designing the IVR and comparison conditions. For example, if a researcher is interested in understanding whether it is the interactivity of an IVR lesson that impacts learning, they should ensure that only *interactivity* is different between the two conditions. However, if they are interested in understanding whether there is a unique benefit of hands-on learning in IVR, the researcher should present the same hands-on activity in both conditions in order to control variables and draw sound conclusions.

K-16 educators

Based on the current research base for IVR in educational settings, our recommendation to instructors interested in using readily available IVR experiences in their own classrooms is to adopt the technology as a motivational tool rather than as a primary learning tool, at least at this point in time. Educators and students alike should benefit from the aspects of IVR we know work well—that is, IVR is an effective motivational tool for students and can help increase their interest in the learning material. However, more focused research on when and how IVR can benefit learning needs to be conducted before it is adopted as the *primary method* of teaching foundational content. One way to implement IVR to leverage its benefits for motivational and affective components of learning is to have students take a "virtual field trip" (i.e., have students experience a location and/or experience they would otherwise not be able to access by using IVR devices) to spark their interest and then leverage more traditional methods of instruction to teach specific content.

VR developers

As for the developers of VR educational environments, it is important to be aware of and incorporate findings from research on the cognitive processes of learning. As discussed, many IVR lessons are not designed according to theories of learning. As such, the way in which the material is presented to learners is oftentimes inconsistent with how the brain processes information. By integrating findings from research on effective design principles, developers can create better lessons that align with human cognition. We recommend that developers work more closely with educational researchers (and vice versa) to create educational content that can be more effectively implemented in educational spaces.

Conclusion

Immersive virtual reality is quite effective in increasing presence and motivation. These outcomes are a large contribution to learning as an important step in getting students to engage in deeper understanding by motivating them to want to learn (Mayer, 2022). However, when and how IVR is effective for student learning has not been well established, with a mixture of studies showing IVR lessons to be better than, equal to, or worse than other modes of instruction. With these inconsistent findings and design/implementation variation that exist in the literature, providing specific implementation guidance to instructors remains difficult. By outlining research gaps that, if filled, may help to explain inconsistent results in the literature surrounding IVR's effectiveness for learning and by providing future research recommendations for researchers, it is our hope that technological tools will be more effectively and appropriately researched and integrated into K-16 classrooms.

References

Akgün, M., & Atici, B. (2022). The effects of immersive virtual reality environments on students' academic achievement: A meta-analytical and meta-thematic study. *Participatory Educational Research, 9*(3), 111–131. https://doi.org/10.17275/per.22.57.9.3

Albert, W. D., Hanson, J. L., Skinner, A. T., Dodge, K. A., Steinberg, L., Deater-Deckard, K., Bornstein, M. H., & Lansford, J. E. (2020). Individual differences in executive function partially explain the socioeconomic gradient in middle-school academic achievement. *Developmental Science, 23*(5), e12937. https://doi.org/10.1111/desc.12937

Alhalabi, W. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behavior & Information Technology, 35*(11), 919–925. https://doi.org/10.1080/0144929x.2016.1212931

Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology, 26*, 2140. https://doi.org/10.25304/rlt.v26.2140

Alrehaili, E. A., & Al Osman, H. (2022). A virtual reality role-playing serious game for experiential learning. *Interactive Learning Environments, 30*(5), 922–935. https://doi.org/10.1080/10494820.2019.1703008

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives*. Longman.

Association of American Universities (2017). *Progress toward achieving systemic change: A five-year status report on the AAU undergraduate STEM education initiative*. https://www.aau.edu/sites/default/files/AAU-Files/STEM-Education-Initiative/STEM-Status-Report.pdf

Baceviciute, S., Terkildsen, T., & Makransky, G. (2021). Remediating learning from non-immersive to immersive media: Using EEG to investigate the effects of environmental embeddedness on reading in virtual reality. *Computers & Education, 164,* 104122. https://doi.org/10.1016/j.compedu.2020.104122

Barnidge, M., Sherrill, L. A., Kim, B., Cooks, E., Deavours, D., Viehouser, M., Broussard, R., & Zhang, J. (2022). The effects of virtual reality news on learning about climate change. *Mass Communication and Society, 25*(1), 1–24. https://doi.org/10.1080/15205436.2021.1925300

Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. David McKay Co. Inc.

Bukoski, R. J. (2019). *The impact of traditional and video modalities on concept retention*. [Doctoral dissertation, Indiana University of Pennsylvania]. ProQuest Dissertations Publishing.

Campos, E., Hidrogo, I., & Zavala, G. (2022). Impact of virtual reality use on the teaching and learning of vectors. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.965640

Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics, 21*(4), 529–538. https://doi.org/10.1109/TVCG.2015.2391853

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Psychology Review, 36*, 100452. https://doi.org/10.1016/j.edurev.2022.100452

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Pearson.

Davis, A., Goulding, M., & Suggate, J. (2017). *Mathematical knowledge for primary teachers* (5th ed.). Routledge. https://books.google.com/s?hl=en&lr=&id=7T8lDwAAQBAJ&oi=fnd&pg=PP1&dq=Mathematical+knowledge+for+primary+teachers&ots=tJtb9cLwj_&sig=uJC1mIqK756wnnVoFvBTg35kvQA#v=onepage&q=Mathematical%20knowledge%20for%20primary%20teachers&f=false

deWinstanley, P. A., & Bjork, R. A. (2002). Successful lecturing: Presenting information in ways that engage effective processing. *New Directions for Teaching and Learning, 89*, 19–31. https://doi.org/10.1002/tl.44

Drake, E. W. (2022). *The impact of virtual reality and the virtual reality learning environment of student assessment outcomes in a unit of study in middle school life science: A quasi-experimental study.* [Doctoral dissertation, The University of Arizona Global Campus]. ProQuest Dissertations Publishing.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. http://doi.org/10.1177/1529100612453266

Ekstrand, C., Jamal, A., Nguyen, R., Kudryk, A., Mann, J., & Mendez, I. (2018). Immersive and interactive virtual reality to improve learning and retention of neuroanatomy in medical students: A randomized controlled study. *CMAJ Open, 6*(1), E103-E109. https://doi.org/10.9778/cmajo.20170110

Fenesi, B., Sana, F., Kim, J. A., & Shore, D. I. (2015). Reconceptualizing working memory in educational research. *Educational Psychology Review, 27*, 333–351. https://doi.org/10.1007/s10648-014-9286-y

Flavia Di Natale, A., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology, 51*(6), 2006–2033. https://doi.org/10.1111/bjet.13030

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Greenwald, S. W., Corning, W., Funk, M., & Maes, P. (2018). Comparing learning in virtual reality with learning on a 2D screen using electrostatics activities. *Journal of Universal Computer Science, 24*(2), 220–245. https://doi.org/10.3217/jucs-024-02-0220

Grenell, A., & Carlson, S. M. (2021). Individual differences in executive function and learning: The role of knowledge type and conflict with prior knowledge. *Journal of Experimental Child Psychology, 206*, 105079. https://doi.org/10.1016/j.jecp.2020.105079

Hassenfeldt, C., Jacques, J., & Baggili, I. (2020). Exploring the learning efficacy of digital forensics concepts and bagging and tagging of digital devices in immersive virtual reality. *Proceedings of the Twentieth Annual DFRWS USA, 33,* 301011. https://doi.org/10.1016/j.fsidi.2020.301011

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry-learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Huang, C. L., Luo, Y. F., Yang, S. C., Lu, C. M., & Chen, A-S. (2020). Influence of students' learning style, sense of presence, and cognitive load on learning outcomes in an immersive virtual reality learning environment. *Journal of Educational Computing Research, 58*(3), 596–615. https://doi.org/10.1177/0735633119867422

Kim, S., Heo, R., Chung Y., Kim, J. M., Kwon, M. P., Seo, S. C., Park, G., & Kim, M. (2019). Virtual reality visualization model (VRVM) of the Tricarboxylic Acid (TCA) Cycle of carbohydrate metabolism for medical biochemistry education. *Journal of Science Education and Technology, 28*, 602–612. https://doi.org/10.1007/s10956-019-09790-y

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), e41097. https://doi.org/10.2196/41097

Klahr, D. (2013). What do we mean? On the importance of not abandoning scientific rigor when talking about science education. *PNAS, 110*(3), 14075–14080. https://doi.org/10.1073/pnas.1212738110

Klingenberg, S., Jørgensen, M. L. M., 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 Psychology, 51*(6), 2115-2138. https://doi.org/10.1111/bjet.13029

Kozhevnikov, M., Gurlitt, J., & Kozhevnikov, M. (2013). Learning relative motion concepts in immersive and non-immersive virtual environments. *Journal of Science Education and Technology, 22*(6), 952–962. https://doi.org/10.1007/s10956-013-9441-0

Lamb, R., Antonenko, P., Etopio, E., & Seccia, A. (2018). Comparison of virtual reality and hands on activities in science education via functional near infrared spectroscopy. *Computers & Education, 124,* 14–26. https://doi.org/10.1016/j.compedu.2018.05.014

Lawson, A. P., & Mayer, R. E. (under review). *Individual differences in executive function affect learning with immersive virtual reality.*

Liu, Q., Tang, Q., & Wang, Y. (2021). The effects of pretraining intervention in immersive embodied virtual reality cardiopulmonary resuscitation training. *Behaviour and Information Technology, 40*(12), 1265–1277. https://doi.org/10.1080/0144929X.2021.1960606

Liu, R., Wang, L., Koszalka, T. A., & Wan, K. (2022). Effects of immersive virtual reality classrooms on students' academic achievement, motivation and cognitive load in science lessons. *Journal of Computer Assisted Learning, 38*(5), 1422–1433. https://doi.org/10.1111/jcal.12688

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

Lui, A. L. C., Not, C., & Wong, G. K. W. (2023). Theory-based learning design with immersive virtual reality in science education: A systematic review. *Journal of Science Education and Technology, 32*(3), 390–432. https://doi.org/10.1007/s10956-023-10035-2

Luo, H., Li, G., Feng, Q., Yang, Y., & Zuo, M. (2021). Virtual reality in K-12 and higher education: A systematic review of the literature from 2000 to 2019. *Journal of Computer Assisted Learning, 37*(3), 887–901. https://doi.org/10.1111/jcal.12538

Madden, J., Pandita, S., Schuldt, J. P., Kim, B., Won, A. S., & Holmes, N. G. (2020). Ready student one: Exploring the predictors of student learning in virtual reality. *PLoS One, 15*(3), e0229788. https://doi.org/10.1371/journal.pone.0229788

Makransky, G. (2020). The immersion principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed). Cambridge University Press.

Makransky, G., Andreasen, 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., Borre-Gude, S., & Mayer, R. E. (2019a). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning, 35*(6), 691-707. https://doi.org/10.1111/jcal.12375

Makransky, G., & Mayer, R. E. (2022). Benefits of taking a virtual field trip in immersive virtual reality: Evidence for the immersion principle in multimedia learning. *Educational Psychology Review, 34,* 1771–1798. https://doi.org/10.1007/s10648-022-09675-4

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*, 937–958. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Petersen, G. B., & Klingenberg, S. (2020). Can an immersive virtual reality simulation increase students' interest and career aspirations in science? *British Journal of Educational Technology, 51*(6), 2079-2097. https://doi.org/10.1111/bjet.12954

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019b). 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

Martella, A. M., & Schneider, D. W. (in press). A reflection on the current state of active learning research. *Journal of the Scholarship of Teaching and Learning*.

Matovu, H., Ungu, D. A. K., Won, M., Tsai, C., Treagust, D. F., Mocerino, M., & Tasker, R. (2023). Immersive virtual reality for science learning: Design, implementation, and evaluation. *Studies in Science Education, 59*(2), 205–244. https://doi.org/10.1080/03057267.2022.2082680

Mayer, R. E. (2022). Cognitive theory of multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press. https://doi.org/10.1017/cbo9781139547369.005

Mayer, R. E., Makransky, G., & Parong, J. (2022). 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

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*(1), 165–173. https://doi.org/10.1037/0022-0663.96.1.165

Office of Educational Technology. (2017, January). *Reimaging the role of technology in education: 2017 national education technology plan update.* U.S. Department of Education. https://tech.ed.gov/files/2017/01/NETP17.pdf

Paas, F., & Sweller, J. (2022). Implications of cognitive load theory for multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 73–81). Cambridge University Press.

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

Parong, J., & Mayer, R. E. (2021a). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning, 37*(1), 226–241. https://doi.org/10.1111/jcal.12482

Parong, J. & Mayer, R. E. (2021b). Learning about history in immersive virtual reality: Does immersion facilitate learning? *Educational Technology Research and Development, 69,* 1433–1451. https://doi.org/10.1007/s11423-021-09999-y

Petersen, G. B., Klingenberg, S., & Makransky, G. (2022). Pipetting in virtual reality can predict real-life pipetting performance. *Technology, Mind, and Behavior, 3*(3), 1–16. https://doi.org/10.1037/tmb0000076

Porter, C. D., Brown, J. R., Smith, J. R., Stager, E. M., Nieberding, M., Ayers, A., & Orban, C. (2019). A controlled study of virtual reality in first-year magnetostatics. *2019 PERC Proceedings.* https://doi.org/10.111+/perc.2019.pr.Porter

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

Sanzana, M. R., Abdulrazic, M. O. M., Wong, J. Y., Ng, K. H., & Ghazy, S. (2022). Lecture-based, virtual reality game-based and their combination: Which is better for higher education? *Journal of Applied Research in Higher Education, 14*(4), 1286–1302. https://doi.org/10.1108/JARHE-09-2020-0302

Stranger-Johannessen, E. (2018). Exploring math achievement through gamified virtual reality. In V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, & M. Scheffel (Eds.), *Lifelong technology-enhanced learning* (pp. 613–616). https://doi.org/10.1007/978-3-319-98572-5_57

Su, Y-S., Cheng, H-W., & Lai, C-F. (2022). Study of virtual reality immersive technology enhanced mathematics geometry learning. *Frontiers in Psychology, 13.* https://doi.org/10.3389/fpsyg.2022.760418

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68,* 1–16. https://doi.org/10.1007/s11423-019-09701-3

Tarng, W., Liao, Y-C., & Ou, K-L. (2022). Applications of virtual reality in learning the concepts of special relativity and mass-energy equivalence. *Universe, 8*(12), 618. https://doi.org/10.3390/universe8120618

Taranilla, R. V., Cózar-Gutiérrez, R., González-Calero, J. A., & Cirugeda, I. L. (2022). Strolling through a city of the Roman Empire: An analysis of the potential of virtual reality to teach history in primary education. *Interactive Learning Environments, 30*(4), 608–618. https://doi.org/10.1080/10494820.2019.1674886

Ward-Horner, J., & Sturmey, P. (2010). Component analyses using single-subject experimental designs: A review. *Journal of Applied Behavior Analysis, 43*(4), 685–704. https://doi.org/10.1901/jaba.2010.43-685

Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments, 24*(6), 1319–1333. https://doi.org/10.1080/10494820.2014.994533

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness in 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

Yang, Q-F., Lin, H., Hwang, G-J., Su., P-Y., & Zhao, J-H. (2022). An exploration-based SVVR approach to promote students' chemistry learning effectiveness. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2022.2135106

Yu, Z (2021). A meta-analysis of the effect of virtual reality technology use in education. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.1989466

Copyright: © 2023. Alyssa P. Lawson and Amedee Marchand Martella. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.