Immersive Virtual Reality Increases Liking but Not Learning with a Science Simulation and Generative Learning Strategies Promote Learning in Immersive Virtual Reality

Guido Makransky[1], Niels K. Andreasen[1], Sarune Baceviciute[1], and Richard E. Mayer[2]

[1] Department of Psychology, University of Copenhagen, Copenhagen, Denmark

[2]Psychological and Brain Sciences, University of California Santa Barbara, CA

Abstract

We investigated the instructional effectiveness of using an interactive and immersive virtual reality (IVR) simulation versus a video for teaching scientific knowledge in two between-subject experiments. In Experiment 1, 131 high school students (84 females) used a science simulation that involved forensic analysis of a collected DNA sample in a virtual laboratory environment rendered in IVR or as a video covering the same material. In Experiment 2, 165 high school students (111 females) replicated the experiment with approximately half of each group being asked to engage in the generative learning strategy of enactment after the lesson--i.e., carrying out the learned procedures with concrete manipulatives. Across both experiments, the IVR groups reported significantly higher perceived enjoyment and presence than the video group. However, no significant differences were found between media for procedural knowledge in Experiment 1 and 2, or transfer in Experiment 2. Also, there was no difference in declarative knowledge across media in Experiment 1, and there was a media effect favoring video in Experiment 2 (η_p^2 = 0.028). Enactment lead to significantly better procedural knowledge (η_p^2 = 0.144) and transfer (η_p^2 = 0.088) in the IVR group but not in the video group. In conclusion, learning in IVR is not more effective than learning with video but incorporating generative learning strategies is specifically effective when learning through IVR. The results suggest that the value of IVR for learning science depends on how it is integrated into a classroom lesson.

Keywords: immersive virtual reality, video, media, science simulations, learning strategies, enactment

Educational Impact and Implications Statement

Which is a more effective way to learn about scientific procedures--in an interactive simulated laboratory rendered in immersive virtual reality (IVR) or with the conventional medium of video? This question was investigated in two between-subject experiments involving a total of 296 high school students. In both experiments, students liked learning in IVR more than from video and felt a greater sense of presence, but they did not learn better in IVR than from video. Experiment 2 added evidence that asking students to carry out the procedures with concrete manipulatives improved performance when learning in IVR but not when learning with video. Since IVR is a popular technology among students, as indicated by higher perceived enjoyment in IVR, using it in combination with a generative learning strategy may be beneficial in fostering student motivation and still sustain learning outcomes when compared to other more traditional media.

Immersive Virtual Reality Increases Liking but Not Learning with a Science Simulation and Generative Learning Strategies Promote Learning in Immersive Virtual Reality

Objective and Rationale

The increased availability of technologies for creating Immersive Virtual Reality (IVR) enable novel ways of engaging learners in education and training (Baceviciute et al., 2020; Bonde et al., 2014; Jensen & Konradsen, 2018; Mikropoulos & Natsis, 2011; Parong & Mayer, 2018). The growing demand for IVR-based educational activities has been stimulated by the enthusiasm of proponents and large-scale investments from technology companies (e.g., Freeman, Becker & Cummins, 2017), but there has been inadequate attention to research evidence concerning its actual educational value (Makransky, Terkildsen, & Mayer, 2019b). It is therefore useful to investigate the conditions under which IVR can benefit students' learning, as well as how to properly implement IVR in educational settings.

There are three main objectives of this study. The first objective is to conduct research in real life settings rather than a lab environment. We seek to investigate whether the findings from previous lab-based studies of IVR (Moreno & Mayer, 2002; Makransky et al., 2019b; Meyer et al. 2019; Parong & Mayer, 2018) generalize to realistic educational contexts. To do this, we conducted two experiments with high school students at an honors science camp (Experiment 1) and in a classroom setting (Experiment 2).

The second objective is to investigate the instructional effectiveness of using an interactive IVR simulation compared to a video of an optimal simulation experience for teaching biological principles. The principles involved polymerase chain-reaction and DNA replication in a Crime-Scene Investigation (CSI), and the main outcome measures included scores on tests of procedural and declarative knowledge, ratings of presence, and ratings of enjoyment. This is a media comparison study, as we only varied the media by which the CSI high school science lesson was delivered. In this media comparison experiment, we note that IVR affords interactivity and immersion, which are related to the psychological factors of agency and presence respectively (Johnson-Glenberg, 2019). Thus, in comparing learning in IVR versus learning with video, it is possible that any observed differences are caused by interactivity or immersion or both. This comparison is relevant because previous research on the topic of learning through IVR compared to conventional media has found mixed results related to the cognitive learning outcomes but positive results related to affective outcomes such as presence and perceived enjoyment. Some explanations for the differences in learning outcomes is that the comparison media differs across studies from desktop VR (e.g., Moreno & Mayer, 2002; Makransky et al., 2019), a PowerPoint slide show (Parong & Mayer, 2018), or a video (e.g., Mayer et al., 2019). Each medium provides different affordances related to the instructional methods that are innately part of lessons, and the studies differ in terms of the outcome measures that are used (Makransky, Borre-Gude & Mayer, 2019). Furthermore, many of these studies have been conducted in a lab setting. Therefore, we conducted experiments in a realistic educational setting that compared learning in IVR versus with a video containing the same information (i.e., Experiments 1 and 2).

The third objective (which is examined in Experiment 2) is to use a value-added study to investigate the effectiveness of having students engage in the generative learning strategy of enactment (Fiorella & Mayer, 2015, 2016) in conjunction with the lesson. This issue is relevant because there is limited research investigating how to optimally integrate IVR in a classroom setting (Meyer et al., 2019). Several studies suggest that adding a generative learning strategy as an instructional method, in combination with the respective media can help learners understand the material in deeper ways (Fiorella & Mayer, 2012, 2016; Parong & Mayer, 2018; Pilegard & Mayer, 2016). Specifically, enactment, which involves asking the learner to engage in relevant physical activity with concrete objects during learning, has proven to be effective for improving learning (Fiorella & Mayer, 2016). Since embodied interactions is one of the defining affordances of IVR (Shin, 2017), adding enactment as a generative learning strategy in combination with IVR might therefore provide a better understanding of the learning potential of IVR.

Definition of Virtual Reality

There are many different ways to define VR. In this study, we take the view that VR is a complex media system, that encompasses a specific technological setup for sensory immersion and is capable of giving users the experience of being in the simulated world (Mikropoulos & Natsis, 2011). VR can be accessed through various displays, such as a desktop VR, VR with a head-mounted display (HMD), or a cave automatic virtual environment (CAVE; Buttussi & Chittaro, 2018). Desktop VR uses a 3-D image on a computer screen (Lee & Wong, 2014). VR with a HMD, which is used in the current study, is attained with a head-mounted display (HMD), that portrays the virtual environment by locating the user's head orientation and position from a tracking system (Makransky & Lilleholt, 2018). A CAVE system is a room where all the walls, as well as the floor, are projection screens (Freina & Ott, 2015). A clear distinction between different types of VR outlined above is the degree of immersion (Cummings & Bailenson, 2016; Makransky & Lilleholt, 2018), which is the vividness offered by the system, and the extent to which the used display system is capable to shut out the outside world (Cummings & Bailenson, 2016). IVR systems offer head and position tracking to various degrees and HMDs render a different image for each eye, creating visual cues for depth perception and a realistic first person perspective. Furthermore, IVR systems increases the size of the visual field of view compared to a monitor. Consequently, desktop VR is considered low immersion, whereas VR accessed through an HMD or a CAVE are regarded as a high-immersion, because the user is surrounded by the virtual environment. The level of immersion also distinguishes VR from other media such as video.

Another distinguishing element of VR is the level of interactivity that a VR system can afford. Interactivity is a technical feature of the virtual learning tool which is related to the learner's sense of user control and agency (Sawyer et al., 2009). Agency is a learner's degree of freedom and control to perform meaningful actions in the virtual environment (Wardrip-Fruin et al., 2009). Most desktop VR systems can be explored interactively by using either a keyboard, a mouse, a joystick, or a touch screen. Some systems have specialized integrated VR controllers, that offer varying degrees of freedom and different possibilities for haptic interaction (Howard, 2019). The interactive element of IVR--such as being able to move and interact with objects in the simulated world--is particularly relevant as it is a complex multidimensional construct (Ritterfeld et al., 2009) that can both facilitate and impede learning (Song et al., 2014). Consequently, it is evident that HMD VR systems are superior to standard media, such as videos, in terms of the level of interactivity, as well as the level of immersion which they can afford the learner. In the following section we describe relevant research related to learning in IVR. This is followed by a theoretical section that describes how the affordances of interactivity and immersion which differentiate IVR from a video could be related to motivational and learning outcomes.

Research on Learning in Immersive Virtual Reality

Previous research posits that IVR has the potential to increase ratings of presence and enjoyment as well as scores on tests of learning (Makransky, Borre-Gude & Mayer, 2019; Makransky & Lilleholt, 2018; Thisgaard & Makransky, 2017). Experiential forms of education have been promoted since the early 1900s by theorists such as Dewey (1913), who embraced the notion that learners need to interact with the world in order to understand it. While previous research investigating the media effect of learning through IVR has consistently reported that it outperforms traditional media on outcomes such as motivation and presence (e.g., Makransky & Lilleholt, 2018; Parong & Mayer, 2018), there is insufficient evidence for the value of IVR as a platform for knowledge acquisition compared to less immersive media (Makransky et al., 2019b; Parong & Mayer, 2018). While some studies report finding positive results on procedural knowledge acquisition (John, Pop, Day, Ritsos, & Headleand, 2018; Li, Liang, Quigley, Zhao, & Yu, 2017), others report null or negative results (e.g., van Ginkel et al., 2019; Leder, Horlitz, Puschmann, Wittstock, & Schütz, 2019; Sacks, Perlman, & Barak, 2013). Similarly, some studies have compared IVR with less immersive media for teaching of declarative knowledge and found either positive results (e.g., Alhalabi, 2016; Webster, 2016), mixed results (e.g., Moreno & Mayer, 2002; Polcar & Horejsi, 2015), or reported IVR to be significantly worse than the compared media (e.g., Makransky et al., 2019b; Parong & Mayer, 2018). One factor which could determine these differences is that the IVR based interventions differ based on various technical features, since IVR comprises a variety of technical artifacts such as hardware, software, physical manipulation devices, etc. Therefore, integrating across these previous studies based on technological features is difficult to the degree that different IVR technologies were involved; however, the research presented above is limited to interactive IVR based simulations involving head-mounted displays as this is the type of IVR used in the current study. There are also many other relevant factors that could explain the differences in results across previous studies including instructional design factors, content area, learners, comparison condition, or assessment instruments used across studies. Therefore, recent studies have focused on investigating how instructional methods generalize across media in an attempt at isolating media and method factors.

A handful of recent studies have investigated the interaction between media and instructional method in the framework of IVR. Makransky and colleagues (2019b) investigated the redundancy principle in an interactive immersive VR simulation or a desktop VR simulation and found that students were significantly more present and rated a higher perceived enjoyment when using IVR compared to desktop VR. They report no interaction between media and instructional method, and no evidence for the redundancy principle across media. However, students had higher retention scores in the desktop VR condition. They used EEG to track cognitive load and found that students in the IVR condition had higher cognitive load late in the learning session.

Some studies have investigated whether different forms of scaffolding can increase the effectiveness of learning in IVR. Meyer, Omdahl and Makransky (2019) compared a non-interactive IVR biology lesson against a video version of the same lesson in combination with pre-training as an instructional method in a two by two design. They found an interaction between media and method on declarative knowledge and transfer. More specifically, students in the IVR condition scored significantly higher on knowledge and transfer outcomes when they received pre-training prior to the lesson compared to the students who did not get pre-training material, but the differences were not significant in the video condition. Furthermore, the IVR lesson with pre-training outperformed the other three conditions (video with and without pre-training, and VR without pre-training). The authors conclude that pre-training was specifically beneficial for the IVR group because it limited the cognitive load in the lesson which may be higher in IVR compared to video.

Parong and Mayer (2018) conducted two experiments. The first was a media comparison study comparing student learning in IVR compared to a self-directed PowerPoint lesson on a desktop computer. In Experiment 2, students either viewed a segmented VR lesson and produce a written summary after each segment, or viewed the VR lesson continuously. They found that the PowerPoint slide show resulted in better learning than the IVR simulation in Experiment 1. In Experiment 2, they found that students learned more when they were given time to summarize the learnt content after segments of the simulation, compared to students who did not have time to summarize the content. The authors conclude that the use of the generative strategy of summarizing gave them time to reflect over the material, thus creating a process for deeper learning.

Taken together, these studies suggest that IVR simulations afford a high level of presence and enjoyment. However, they may also lead to higher levels of cognitive load and thus a reduction in learning in settings where there is not enough pre-training (Mayer et al., 2019), or when students do not have ample time to reflect over the material that they have learned (Parong & Mayer, 2018). Overall, these studies suggest that students who learn through IVR may particularly benefit from scaffolding.

To understand why this might be the case, in the next section we describe the specific features of IVR systems that differ from standard media and how these features might facilitate and impede learning. The two major differences between learning in IVR compared to a PowerPoint lesson or an instructional video are the levels of interactivity and immersion.

Theoretical Background and Hypothesis

The study's theoretical foundation is based on motivational theories concerning interest (e.g., Pekrun, 2006; Renninger & Hidi, 2016; Schiefele, 2009), cognitive theories, such as the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2009, 2014), the Embodied Cognition framework (Wilson, 2002), and Generative Learning Theory (Fiorella & Mayer, 2015, 2016; Wittrock, 1974, 1989).

Prior research demonstrates that student motivation is important for learning, as motivated students are more engaged in the learning material, show greater persistence when trying to understand the material, and are more resilient when dealing with potential obstacles in understanding (Parong & Mayer, 2018). A reason for using IVR in the classroom is grounded in interest theory (Renninger & Hidi, 2016; Schiefele, 2009), which suggests that students work harder when they are intrinsically interested in the material, or if the lesson itself elicits situational interest in the learner (Brom et al., 2019; Mayer, 2008; Schiefele, 2009; Wigfield, Tonks & Klauda, 2016). A meta-analysis conducted by Schiefele, Krapp and Winteler (1992)**,** found a correlation between students' self-ratings of how interested they were in specific school subjects and how well they did in school overall. IVR learning experiences can elicit situational interest because students perceive them as being more enjoyable than lessons in standard media (e.g., Makransky & Lilleholt, 2018; Parong & Mayer, 2018). The rationale for measuring students' perceived enjoyment is based on the expectation that using IVR for presenting lessons will make learning a fun experience, in which both engagement and motivation are affected (Vogel et al., 2006).

A theoretical explanation for this can be found in Pekrun's (2006) Control Value Theory of Achievement Emotions (CVTA), which distinguishes between achievement emotions pertaining to ongoing achievement-related activities and outcomes. An important activity emotion is enjoyment, which is triggered when the achievement activity is positively valued and controllable. Enjoyment benefits performance because the learner focuses attention on the task, which leads to higher intrinsic and extrinsic student motivation (Pekrun, 2006). IVR has been shown to provide learners with a high level of perceived control (Makransky & Lilleholt, 2018) compared to less immersive media which can lead to a high sense of psychological presence and enjoyment (Makransky & Pedersen, 2019). Positive emotions can serve as mediators between the multimedia lesson and learning outcomes (Makransky & Pedersen, 2019; Pekrun, 2006; Pekrun, Goetz, Frenzel, Barchfeld & Perry, 2011; Pekrun & Stephens, 2010; Reyes, Brackett, Rivers, White, & Salovey, 2012) as learners are more likely to engage in high levels of generative processing (Mayer, 2014). Furthermore, studies have persistently shown that IVR is associated with a higher self-reported perceived enjoyment in learning contexts because they prompt more agency and presence**,** compared to more conventional media (e.g., Buttussi & Chittaro, 2018; Makransky et al., in press; Makransky & Lilleholt, 2018; Makransky et al., 2019; Meyer et al., 2019; Parong & Mayer, 2018). We therefore predict that a lesson presented using IVR will result in significantly higher ratings of perceived enjoyment than a lesson presented using a video (Hypothesis 1 in Experiments 1 and 2).

Interactivity and immersion, which are two obvious distinguishing characteristics of IVR compared to a video, have been repeatedly associated with psychological presence (Makransky & Lilleholt, 2018; Makransky & Pedersen, 2019; Terkildsen & Makransky, 2019). Psychological presence is defined as an overall subjective sensation of "being there" in virtual world to the point where the virtuality of the environment goes unnoticed and feels like the dominant reality (Barfield, Zeltzer, Sheridan, Slater & 1995; Ijsselsteijn, de Ridder, Freeman & Avons, 2000; Lee, 2004; Terkildsen & Makransky, 2019). Psychological presence is relevant for understanding the process of learning in VR (Makransky & Lilleholt, 2018; Makransky & Pedersen, 2019) because it can facilitate learning by causing students to engage in the learning material as intended. For instance, feeling like you are in a real lab, and working with actual relevant procedures, which are a part of a story or a game-like experience have been shown to be engaging for learning (Makransky & Pedersen, 2019). Research has found that IVR lessons lead to higher levels of psychological presence than less immersive media (e.g., Makransky & Lilleholt, 2018; Makransky et al., 2019b). Therefore, we predict that a lesson presented using IVR will result in significantly higher ratings of psychological presence than a lesson presented using a video (Hypothesis 2 in Experiment 1 and 2).

The Cognitive Theory of Multimedia Learning (CTML) proposes several empirically based design principles for multimedia, with the goal of enhancing learning (Mayer, 2014a). CTML differentiates between three types of cognitive processing that take place during learning: *Extraneous processing, essential processing,* and *generative processing* (Mayer, 2014b, p. 60). Extraneous processing occurs any time the instructional goal is not supported by the multimedia including poor instructional designs, usability issues, or distractions during learning. Previous research suggests that learners experience higher amounts of sensory stimuli in IVR when compared to less immersive media, which can lead to overload of extraneous processing and thereby decreased learning (Makransky et al., 2019b; Richards & Taylor, 2015; Slobounov, Ray, Johnson, Slobounov, & Newell, 2015). Essential processing is linked to the complexity of the learning material. Generative processing aims to make sense of the material, and is related to the learner's motivation to exert more effort (Mayer, 2014b). The different cognitive processes are proposed to be additive, meaning that if a learner is engaged with something that requires unnecessary extraneous processing, they might not have enough working memory capacity left for essential and generative processing. An important reason for comparing an IVR simulation with a less immersive media system, i.e. a video, is the assumption that adding IVR to a lesson might create extraneous processing, exceeding the student's ability to engage in cognitive processes aimed at making sense of the material. This extraneous processing can be a consequence of an increased visual field of view, the control interface, unfamiliarity with the technology, or a number of other differences, which could make it more challenging to select, organize, and integrate relevant information (Makransky et al., 2019b).

Several studies have found mixed results when comparing different levels of interactivity, suggesting that interactivity is a multifaceted concept (Chittaro & Sioni, 2015; Ritterfeld et al., 2009; Saywer et al., 2017; Song et al., 2014; Steinmann et al., 2017). Meta-analyses have found that students prefer to navigate through simulations themselves (Vogel et al., 2006), and that high levels of user control result in higher estimates of self-efficacy and transfer (Gegenfurtner et al., 2014). However, allowing learners to freely interact with a learning environment is not always desirable for promoting learning (Mayer, 2004). For instance, Sawyer and colleagues (2009) manipulated the level of agency in a game-based learning environment and found that students in a low agency condition achieved greater learning gains that students who had high agency or no agency. The low agency participants had to navigate through a partially ordered sequence, while the high agency participants could freely navigate the open 3D world, and a no agency participants watched an expert follow an ideal path through the game.

Although increased student agency is related to higher levels of motivation and involvement (Wardrip-Fruin et al., 2009; Snow et al., 2015), it can also lead to undesirable behaviors when learners do not monitor and regulate their cognitive, metacognitive, and affective process during learning (Winnie et al., 2014; Sawyer et al., 2017). Specifically, novice students who are given too much freedom have been found to have difficulty in selecting, organizing and integrating relevant information in discovery learning settings (Mayer, 2004; Kirchner at al., 2006). In IVR these factors are intensified because students feel a high sense of presence (Makransky et al., 2019b; Parong & Mayer, 2018) meaning that their interactions seem real thereby increasing their enjoyment and motivation which are factors beneficial to learning (Makansky & Lilleholt, 2018; Makransky & Petersen, 2019). Nonetheless, such an engaging environment also tempts learners to engage in hedonic activities that have an entertainment value, but are not beneficial to their learning (Moreno & Mayer, 2002; van der Heijden, 2004), and IVR adds complexity, which can distract the learner.

In conclusion, although IVR is intended to lead to higher levels of generative processing by engaging the learner through a higher level of interaction and agency, as well as a higher sense of presence and enjoyment, there is also empirical and theoretical support for why IVR could cause extraneous processing that is detrimental to learning outcomes. Previous media and methods experiments have not found a media effect of IVR on declarative knowledge (e.g., Mayer & Moreno, 2002; Makransky et al., 2019b; Parong & Mayer, 2018). Therefore, although there is theoretical support for the IVR condition to perform significantly better than the video condition on the declarative knowledge test (Hypothesis 3 in Experiment 1 and 2), there is also theoretical and empirical support based on cognitive theories and empirical evidence related to interactivity for the opposite.

Embodied cognition suggests that the way we think and make sense of the world depends greatly on our sensorimotor system and bodily interactions with the surrounding environment (Wilson, 2002). Relevant to learning procedural knowledge in VR, embodied cognition contends that direct bodily manipulation of external representations is imperative to learning, and thereby suggests to capitalize on these in the design and evaluation of learning experiences (Jang, Vitale, Jyung, & Black, 2017). At the same time, it is suggested that in order for physical manipulation to benefit learning, bodily actions have to be carefully coordinated with employed representational features (Jang, Vitale, Jyung, & Black, 2017), and made relevant for a given learning activity (Dalgarno & Lee, 2010). Based on this perspective, learning of procedural knowledge and skills can be increased when the learner performs physical activity that is meaningful to the learning concept and when it has been represented in a way that it guides particular motor actions. User control can be an important factor in gaining procedural knowledge through more direct manipulations, and due to positive results in most previous articles investigating the media effect on procedural knowledge (John, Pop, Day, Ritsos, & Headleand, 2018; Li, Liang, Quigley, Zhao, & Yu, 2017), we predict a main effect on procedural knowledge with the IVR condition outperforming the video condition (Hypothesis 4 in Experiment 1 and 2).

Finally, we predict that the additional enjoyment and presence in IVR will lead to deeper cognitive processing and better transfer test scores than the video (Hypothesis 5 in Experiment 2). However, based on cognitive theories it is also possible that the advantages of IVR are mitigated when the physical interactions are not intuitive or conductive to learning, or when IVR leads to extraneous processing (Makransky et al., 2019b).

Generative learning theory (GLT) is based on Wittrock's (1974, 1989) generative model of learning, hypothesizing that learners are not "passive consumers of information" (1989, p. 348), but actively "generate perceptions and meaning that are consistent with their prior knowledge" (1974, p. 88). This is consistent with theorists such as Dewey (1913) who posited that student learning happens through practical experiences related to real situations and tasks, in which they actively interact with the environment. GLT provides the basis of different generative learning strategies with the intention of promoting generative learning. These learning strategies are presented by Fiorella and Mayer (2016) as, *Summarizing, Mapping, Drawing, Imagining, Self-testing, Self- explaining, Teaching, and Enacting.* The main purpose of adding a generative learning strategy to a lesson, is to stimulate learners to reflect and integrate prior knowledge with the learning material, thereby helping the learner to construct a more meaningful mental representation of the material (Fiorella & Mayer, 2016). This is particularly relevant for IVR, as these lessons can be highly engaging but can cognitively overload learners, thereby limiting their ability to properly reflect and self-regulate during a lesson.

Extending initial work by Parong and Mayer's (2018), the current study seeks to investigate the efficacy of adding a new kind of generative learning strategy, enactment, as an instructional method in combination with a video or an IVR simulation. Enactment involves activity-based instructions during learning, related to and contextualized with the lesson content, making it particularly relevant as a follow-up activity for learning procedures in simulated environments (Fiorella & Mayer, 2016). The benefits of enactment are deeply grounded in the learner's physical interactions with the external world, and it is therefore relevant to use in combination with IVR, as the medium utilizes activity-based content for learning (Bailenson, 2018). In the study by Parong and Mayer (2018), students' declarative knowledge increased when using summarizing as a generative learning strategy. However, enactment relates more to a reflection process based on learning a physical skill which is associated more with procedural knowledge (e.g. conducting a DNA test and pipetting), rather than declarative knowledge (e.g. learning facts about DNA). Therefore, we posit that enactment may help the learner in a different way, as the learning material in the current study is more physically connected to movements (Carbonneau, Marley, & Selig, 2013; Fiorella & Mayer, 2016). Most of the current research investigating the efficacy of enactment has been conducted in combination with a classroom lesson without any multimedia technology (Carbonneau et al., 2013; Cook, Mitchell, & Goldin- Meadow, 2008; Fujimura, 2001; Glenberg, Goldberg, & Zhu, 2011). Due to the fact that IVR is generally considered to be an embodied medium (Riva, 2008), and that prior studies have shown increased procedural knowledge skill acquisition and transfer in IVR (e.g. Makransky et al., in press; Aggarwal et al., 2006) we would expect the generative strategy of enactment to be especially pertinent to this media.

Based on the theory and prior research, we predict that students who engage in the generative learning strategy of enactment will perform better on tests measuring understanding (procedural knowledge and transfer) but not declarative knowledge, as the instructional method relies heavily on performing procedural skills to gain a deeper understanding of the content. Furthermore, we predict that enactment will especially benefit students learning in IVR because it will encourage reflection and self-regulation which may be hampered when learning in a highly engaging, yet cognitively demanding media like IVR (Hypothesis 6 in Experiment 2).

Summary of Theory and Predictions

Based on interest theory (Renninger & Hidi, 2016) theories of emotion (Pekrun, 2006), we expect that learning in immersive virtual reality is more enjoyable and creates greater presence than learning with conventional media, and this boost in affective reaction primes better learning outcomes. Therefore, we test several basic hypotheses:

1. Students who learn in IVR will report higher levels of enjoyment than students who learn with video (Experiments 1 and 2).

2. Students who learn in IVR will report higher levels of presence than students who learn with video (Experiments 1 and 2).

3. Students who learn in IVR will score higher on tests of declarative knowledge than students who learn with video (Experiments 1 and 2).

4. Students who learn in IVR will score higher on tests of procedural knowledge than students who learn with video (Experiments 1 and 2).

5. Students who learn in IVR will score higher on tests of transfer than students who learn with video (Experiment 2).

In contrast, if an increase in affective reaction and an increase in visual complexity creates distraction as suggested by the cognitive theory of multimedia learning (Mayer, 2009, 2014a) and cognitive load theory (Sweller, Ayres, and Kalyuga, 2011), then hypotheses 3, 4, and 5 would not predict that the IVR group performs better on learning outcome tests.

Finally, based on generative learning theory (Fiorella & Mayer, 2015, 1016; Wittrock, 1974, 1979), as summarized in hypothesis 6, we expect enactment to improve learning outcome performance, particularly when learning in IVR, which may otherwise cause distraction.

6. Students who are required to engage in the generative learning strategy of enactment will perform better on learning outcome tests than those who do not, especially on tests measuring understanding (such as procedural knowledge and transfer) and especially when learning in IVR (Experiment 2).

Experiment 1

Participants and Procedure

The sample consisted of 131 high school students (47 boys and 84 girls) between the ages of 17 and 20. A sensitivity power analysis provided an estimate of the minimum detectable effect size for this sample size with 80% power. Using GPower (Faul et al., 2009) we conclude that effect size has to be at least d = 0.21 (small effect size) for an alpha of .05, meaning that the sample size was sufficient for detecting a relatively small effect size. The experiment was part of a larger study that took place during three one-week long honors science camps, where students participated in different mandatory workshops including a VR-workshop, which constituted this study. Students had been pre-selected to take part in the science camp based on being outstanding students within the natural sciences at their respective high schools. Each camp consisted of between 40 and 50 students and had identical setups. Prior to partaking in the lesson, all students were gathered in a lecture hall where they received randomized ID numbers and completed the pre-test, which included demographic characteristics and a prior knowledge survey.

Students were then divided into the video condition (N = 67) or the IVR condition (N = 64) based on the random ID number they had been assigned. Then, the groups were separated into two different classrooms. While the students in the video condition just watched the video, the students in the IVR condition were divided into groups of roughly 10 and additionally received a five-minute oral introduction on how to use the HMD and navigate in the virtual laboratory before entering the simulation. A total of 21 Samsung Gear VR headsets with matching phones had been set up, and students used the IVR-simulations individually with approximately 10 students using the IVR-simulation at a time. Therefore, some students in the IVR condition who had to wait roughly 15 minutes before using the simulation stayed under supervision until they could start. Students were told that they could take off the HMD at any time if they felt discomfort and that they could stop the session at any time for whatever reason without any penalty. None of the students stopped the session, but a few students mentioned that the HMD was not comfortable to have on in post-intervention interviews.

There were four research assistants available in the IVR classroom, and two in the video classroom to supervise the sessions. Immediately after the learning experience, students in both groups were given the post-test. Students in the video group were given the test in the same classroom, and students in the IVR group took the test in an adjacent classroom as to not be distracted by the students still using the simulation. The post-test included scales to investigate students' self-reported perceived enjoyment and presence, as well as tests of declarative and procedural knowledge. Both experiments followed guidelines for research with human subjects and received approval from the institutional ethics committee.

Materials

The multimedia lesson. The learning intervention consisted of either a VR simulation "Polymerase Chain Reaction Virtual Lab" developed by Labster (Labster, 2019), or a high-quality video recording of an optimal run through the simulation (see Figure 1). The simulation is a gamified learning experience that revolves around the narrative of a crime-scene investigation involving forensic analysis of the collected DNA sample in a realistic laboratory environment and supplementary animations of micro-level biological processes such as DNA replication. A science lesson about this subject was chosen because it utilizes the affordances associated with IVR, as it allows the learner to interact with expensive equipment in a safe and controlled environment as well as have access to an immersive experience into micro-biological processes that are not always visible to the naked eye.

In the simulation, the learner is a forensics expert. Finding themselves at a crime-scene, learners must find biological evidence, and further analyze the material in a real scientific laboratory to find the suspected murderer. During the simulation, information is provided through a pedagogical agent, a female lab assistant, who narrates and guides the learner through the simulation (Makransky, Wismer, & Mayer, 2018). The interactivity in the simulation occurred through movements of the head, allowing the learner to control where they focus their attention in the 360-degree virtual environment, at their own pace. Furthermore, students could actively interact with items in the lab using a touch and click system with the control pad located on the right side of the HMD. This allowed students to use a pipette to prepare laboratory samples and allowed them to interact with other objects on a science workbench by clicking on the control pad. Although the simulation provides opportunities for interaction, agency is limited as students are constrained to conduct laboratory activities in a fixed linear sequence, where they are required to complete specified sets of problem-solving actions before moving to the next task. Each set of tasks is guided and supported by the pedagogical agent. Furthermore, learners are asked to make responses to gamified multiple-choice questions, which functions as a retrieval practice activity (Makransky, et al., 2019a). The IVR simulation ran on a Samsung Galaxy S7 or S8 phone inserted into a Samsung Gear VR HMD that features binocular vision. Headphones were used for auditory output (see left panel of Figure 2). The system has a field of view of 101˚, with an interpupilary distance of 62mm (fixed), and eye relief of 10mm. Head tracking is computed using sensorfusion of accelerometer, gyro, and proximity sensors. The simulation ran at 46.7 frames per second. To reduce the likelihood of motion sickness, the immersed participant would not move around physically or in VR. Each participant was given their own HMD with instructions on how to put it on and use it. The simulation took between 10 and 20 minutes to complete.

For the video condition the simulation was screen-captured, creating a video recording. An essential consideration in the process of making the video was recording all of the relevant information in order to make sure that participants learning through this format had access to similar visual information. The video is therefore a recording of an optimal experience of the VR simulation that contained all the procedural steps and conceptual information necessary for learning the intended lab protocols. The video was presented on a large screen that was easily visible by the entire group of students and took 14.56 minutes (see right panel of Figure 2).

Pre- and post-questionnaires. The pre-questionnaire consisted of a prior knowledge survey and demographic characteristics. The prior knowledge survey contained one question asking students to rate their knowledge of DNA replication and synthesis on a scale from 0 (very low) to 4 (very high), and seven questions asking students if they had specific knowledge related to the lesson (e.g. “Do you know what a PCR-machine is?”) where students could respond 0 (no), or 1 (yes). The scale had a Cronbach’s alpha reliability of 0.76.

The post-questionnaire consisted of a perceived enjoyment and presence scale as well as a declarative knowledge and procedural knowledge test. The perceived enjoyment scale was adapted from Tokel and Isler (2013) and consisted of three items (e.g. “I find using this kind of simulation enjoyable”). The average item score was used as an outcome variable for this scale, which had a Cronbach’s alpha reliability of 0.88. The Multidimensional Presence Scale for Virtual Reality Environments (Makransky, Lilleholt & Aaby, 2017) was used to measure presence. The scale includes 15 items, including five items each to assess the three sub-dimensions of environmental (e.g., “The virtual environment seemed real to me”, social (e.g., “I felt like I was in the presence of another person in the virtual environment”), and self-presence (e.g., “I felt like my virtual embodiment was an extension of my real body within the virtual environment). Although each scale had acceptable Cronbach’s alpha coefficients (between 0.84 and 0.93) only the average item score across all 15 items (alpha = 0.93) is used in this study because the results were consistent across all sub-scales. The perceived enjoyment and presence scales were rated on a five-point Likert scale ranging from (1) strongly disagree to (5) strongly agree.

The declarative knowledge test included 19 multiple-choice questions and included conceptual and factual knowledge questions related to the information presented in the simulation or video (e.g. “What happens if you use the same set of primers on two different DNA tests? The product of the polymerase chain reaction will have: A) random DNA sequences, B) different lengths, C) identical DNA sequences, D) the same lengths"). Students received one point for each correct question and the test had a Cronbach’s alpha of 0.70.

The procedural knowledge test included three open-ended questions (e.g. “Describe in steps how to use a pipette to prepare laboratory samples. Mention as many steps as possible."). For the open-ended questions, students received one point for each correct idea that was mentioned and could score up to five points for the first question, up to eight points for the second question, and up to five points for the third question, yielding a maximum total score of 18. The questions were scored by two trained raters who used a rubric and were unaware of each respondent's treatment group. The raters' scores had a correlation of 0.74. The 18 items had a Cronbach's alpha reliability of 0.78 and 0.68 for the raters respectively. Since the raters did not completely agree, disagreements were resolved by consensus in a meeting to obtain a final common score.

Results

Do the groups differ on basic characteristics? There was no difference between the IVR (*M* = 7.63, *SD* = 2.10) and the video (*M* = 7.73, *SD* = 2.25) condition on the prior knowledge scale ($t_{(129)} = .280$; $p = .780$). Furthermore, a chi-square test indicated that the groups did not differ significantly in the proportion of males and females, $\chi 2$ (1, *N* = 131) = 1.165, $p = .280$. Therefore, we conclude that the groups did not differ on these basic characteristics prior to the experiment.

Do the groups differ on enjoyment and presence ratings? The top two rows of Table 1 present the mean scores and standard deviations of the groups on enjoyment and presence ratings, and related statistics. Consistent with hypothesis 1, as shown in the first row of Table 1, the IVR group (*M* = 4.22, *SD* = 0.56) had significantly higher perceived enjoyment scores than the video group (*M* = 3.67, *SD* = 0.84), $t_{(129)} = 4.407$; $p < .001$, $d = 0.79$. Consistent with hypothesis 2, as shown in the second row of Table 1, the IVR group (*M* = 3.64, *SD* = 0.48) had a significantly higher presence rating than the video group (*M* = 2.92, *SD* = 0.66), $t_{(129)} = 7.127$; $p < .001$, $d = 1.26$. We conclude that IVR was more enjoyable and created greater sense of presence than video.

Do the groups differ on learning outcomes? The bottom two rows of Table 1 present the mean scores and standard deviations of the groups on the declarative knowledge and procedural knowledge tests. In contrast to hypothesis 3, as shown in the third row of Table 1, the IVR group (*M* = 14.31, *SD* = 2.93) did not score significantly higher than the video group on the declarative knowledge test (*M* = 14.35, *SD* = 2.59), $t_{(129)}$ = -.015; *p* = .925, *d* = 0.01. In contrast to hypothesis 4, as shown in the bottom row of Table 1, the difference between the IVR group (*M* = 10.47, *SD* = 2.79) and the video group (*M* = 10.07, *SD* = 2.58) on the procedural knowledge test did not reach statistical significance, $t_{(129)}$ = .841, *p* = .402, *d* = 0.15.

In conclusion, Experiment 1 provided evidence for the affective benefits associated with an IVR lesson compared to a video, but there was no evidence of the learning benefits of IVR compared to the video.

Discussion and Limitations of Experiment 1

The results from Experiment 1 are consistent with previous literature that has found that IVR leads to improved affective outcomes including perceived enjoyment and presence, but no differences on the learning outcomes involving declarative knowledge (e.g., Makransky et al., 2019b; Parong & Mayer, 2018). The results related to procedural knowledge acquisition were surprising, however, given that some previous studies have found a positive media effect with IVR outperforming less immersive media on procedural knowledge tests (e.g, John, Pop, Day, Ritsos, & Headleand, 2018; Li, Liang, Quigley, Zhao, & Yu, 2017). The results suggest that a higher level of interactivity and immersion that are unique affordances of IVR compared to the video did not result in a deeper understanding of the procedures. One possible explanation for the results could be that the use of the IVR simulation resulted in higher cognitive load than the video. This is supported by previous IVR research (Makransky et al., 2019b; Meyer et al., 2019; Parong & Mayer, 2018) as well as multimedia research (Adams et al., 2012; Schrader & Bastiaens, 2012) that has suggested that more engaging or immersive learning environments do not always lead to better learning outcomes due to higher cognitive load. Extraneous cognitive load could be a consequence of a lack of interaction fidelity, but could also arise from the complexity of the virtual environment. This could occur if learners do not attend to important information in their self-directed learning experience because they have too many things that they can attend to (Makransky et al., 2019b).

Regarding cognitive load arising from the virtual environment, Meyer and colleagues (2019), found that an IVR lesson only outperformed a video lesson when cognitive load had been limited by introducing students to fundamental concepts in a pre-training session. Furthermore, Parong and Mayer (2018) showed that the introduction of a generative learning strategy within an IVR lesson could significantly improve performance on learning. The novelty effect of IVR could also increase entertainment value in addition to cognitive load due to unfamiliarity. This hypothesis could be feasible, as only 5% of the students in the IVR group had used IVR for more than 2 hours before the experiment, 36% had used it but for less than 2 hours in total, and 59% had never used IVR. A further possible explanation could be that the students already knew most of the material presented in the lesson, as students reported having participated in real laboratory experiments on the topics covered in the lesson. This was the case because this sample of students was at the science camp because of their high level of interest and outstanding performance in natural science.

Experiment 2

As interest in IVR has increased over the last few years, some research suggests that technology improves learning outcomes when it is appropriately implemented in classrooms based on appropriate instructional methods (Abrahamson, Sánchez-García, & Abrahamson, 2016). Building on a pilot study (Andreasen, 2019), we wanted to conduct a well-controlled experiment: (a) with a sample of high school students that would not be as familiar with the content of the lesson; (b) where we introduced a generative learning strategy in conjunction with the lesson in both conditions; and (c) where we further investigate students' knowledge transfer to see how well they can translate their newly acquired knowledge into other situations and contexts.

Participants and Procedure

The sample consisted of 165 high school students (111 females and 54 males) at three different high schools, who all gave consent to participate. We performed a sensitivity power analysis using GPower (Faul et al., 2009) for an ANCOVA with fixed effects, main effects and interactions. The results show that the effect size has to be at least d = 0.22 (small effect size) for an alpha of .05, indicating that the sample size was sufficient for detecting a relatively small effect size. The participants were recruited by contacting teachers and schools, who showed interest in the experiment and were willing to participate during class time. Initially, participants were given a pretest, which included the same prior knowledge survey as in Experiment 1 and demographic characteristics (see Figure 3 for an overview of the steps used in the experiment). Then, they were randomly assigned to one of the four conditions: IVR (n = 42), IVR with enactment (n = 41), video (n = 39), and video with enactment (n = 43). Each condition took place in separate classrooms. This was done deliberately, to keep the IVR and video conditions separated and to prevent participants from distracting each other. All conditions started with students either experiencing the VR simulation or watching the video. The IVR was accessed through an identical HMD system as in Experiment 1 (see left panel of Figure 4). Students in the IVR condition were told that they could take of the HMD at any time, but as in Experiment 1, none of the students experienced symptoms that caused them to stop the session. The students in the video condition viewed the video on their own laptop computer with earphones to prevent them from being distracted (see right panel of Figure 4).

Following the intervention, groups without enactment immediately took a posttest that included the same perceived enjoyment (Cronbach’s alpha = 0.90), presence (Cronbach’s alpha = 0.94), and procedural knowledge measures as in Experiment 1. In order to increase reliability, the declarative knowledge test from Experiment 1 was modified slightly by changing the wording of some questions and removing and replacing other questions. Therefore, the final declarative knowledge test used in Experiment 2 included 20 items, which covered the same content, but is not directly comparable to the test used in Experiment 1. The test had a Cronbach’s alpha reliability of 0.79. The procedural knowledge test was the same as in Experiment 1 and was scored in the same way yielding a maximum total score of 18. The questions were marked blindly by the same two raters as in Experiment 1, and their independent scores had a correlation of 0.90. The 18 items had a Cronbach's alpha reliability of 0.90 and 0.91 for the raters respectively. Since the ratings did not completely match, disagreements were resolved in a meeting to obtain a final common score. The transfer test included a case-question that assessed the students' ability to apply the learned knowledge to a novel situation "In 2014, a two-months old girl-baby was accidentally separated from her parents in a park. The parents immediately reported this to the police, so the police could search for their missing child. A few years later, the police found an orphan girl in an orphanage that the parents claimed to be their daughter. Based on you knowledge of PCR and gel electrophoresis, do you think these techniques could be used to confirm if the girl is their daughter or not? Describe exactly how you would confirm that". The test was designed to measure how well participants were able to use knowledge from the lesson in a different context. Participants' answers were scored from zero to two based on a scoring rubric. The transfer test was marked blindly by the same two raters as the procedural knowledge test, and their independent scores had a correlation of 0.81. Disagreements were resolved in a meeting to obtain a final common score.

Enactment took place outside of IVR with physical props. The groups with enactment were instructed to manipulate with props on a table that resembled all the laboratory tools and equipment that were present in the virtual laboratory after the IVR/video (see left panel of Figure 5). The enactment exercise was divided into three parts that students had to repeat for a period of two minutes each (a total of six minutes). In the first part, students were instructed to enact the steps required to make a DNA sample ready for the PCR-machine (the machine with the color red). In the second part, students were instructed to enact the required steps to make the DNA sample ready for the gel-electrophoresis (the machine with the color green). Finally, in the third step students were instructed to enact the entire procedure from beginning to end. Students were told to enact the exact procedure with the presented props, the same way they remember doing it in the IVR simulation or the video and to continue doing so until the time was up. Students were set up so that they could not see each other, and they were not able to communicate, and were not given any other hints or instructions. When the enactment drill was over, the students took the same post-test as the other groups.

The enactment material used in this study consisted of homemade props, with the purpose of recreating the laboratorial surroundings as found in the simulation (see right panel of Figure 5). The props consisted of all the machines and objects needed to make a polymerase chain reaction (PCR) analysis, which the students were already familiar with from watching the video or using the IVR simulation.

Results

Do the groups differ on basic characteristics? Before investigating the research questions, we investigated if the groups differed on gender and prior knowledge. A one-way ANOVA indicated that the groups did not differ significantly on prior knowledge, ($F_{(3,161)}$ = 1.004, p = .393). However, a chi-square test indicated that the groups differed significantly in the proportion of boys and girls, X^2 (N = 165) = 11.081, p = .011. In conclusion, there was no evidence of differences between the groups on prior knowledge, but a significant imbalance in gender distribution existed before the start of the experiment. Therefore, all subsequent analyses are performed by including gender as a covariate.

Do the groups differ on enjoyment and presence ratings? The top two rows of Table 2 show the means and standard deviations of the groups on enjoyment and presence ratings, along with corresponding statistics. Hypotheses 1 and 2 were investigated with two factorial ANCOVAs with media (IVR vs. video) and method (enactment vs no enactment) as independent variables, and perceived enjoyment and presence as dependent variable respectively, and gender as a covariate. Consistent with hypothesis 1, as shown in the first row of Table 2 there was a significant main effect for media with the IVR groups reporting higher perceived enjoyment (M = 3.96, SD = 0.85) than the video groups, (M = 3.64, SD = 0.88), $F_{(1,159)}$ = 5.571, p = .047, η_p^2 = 0.025. The main effect for method, $F_{(1,159)}$ = 1.012, p = .316, and the interaction between media and method, $F_{(1,159)}$ = .002, p = .961 were not significant. Therefore, as in Experiment 1, we conclude that the students feel significantly more enjoyment when learning through an IVR simulation as compared to learning with video.

Consistent with hypothesis 2, the second row of Table 2 shows a significant main effect for media with the IVR groups reporting higher presence (M = 3.07, SD = 0.76) as compared to the video groups (M = 2.42, SD = 0.75), $F_{(1,159)}$ = 29.513, p < .001, $\boldsymbol{\eta_p^2 = 0.157}$. The main effect for method, $F_{(1,159)}$ = 1.023, p = .313, and the interaction between media and method, $F_{(1,159)}$ = .166, p = .684, were not significant. Therefore, we conclude that the students feel significantly more presence when learning through an IVR simulation as compared to learning with a video.

Do the groups differ on learning outcomes? The bottom three rows of Table 2 show the means and standard deviations of the groups on the declarative knowledge test, procedural knowledge test, and transfer test, respectively. Hypotheses were investigated with factorial ANCOVAs with media (IVR vs. video) and method (enactment vs no enactment) as independent variables, and declarative knowledge, procedural knowledge and transfer as dependent variable respectively, with gender as a covariate.

In contrast to hypothesis 3, as summarized in the third line of Table 2, there was a significant main effect for media showing that the groups that used the video (M = 10.55, SD = 4.25) scored significantly higher than the VR groups (M = 9.60, SD = 3.97) on the declarative knowledge test, $F_{(1,160)}$ = 4.559, p = .034, η_p^2 = 0.028. However, the main effect for method $F_{(1,160)}$ = .002, p = .964 and the interaction between media and method $F_{(1,160)}$ = .365, p = .547 were not significant. We conclude that students learned more declarative knowledge from the video than from the IVR simulation independent of enactment, and that enactment did not contribute to students' declarative knowledge acquisition, as the enactment drill primarily focused on procedural knowledge.

In contrast to hypothesis 4, as summarized in the fourth line of Table 2, there was no significant effect of media indicating that the VR groups (*M* = 6.99, *SD* = 4.85) did not differ significantly from the video groups (*M* = 7.78, *SD* = 4.80) on the procedural knowledge test, $F_{(1,160)} = 2.005$, $p = .159$. However, there was a significant interaction between media and method on the procedural knowledge test indicating that enactment had a greater effect when learning in IVR than by video as expected, $F_{(1,160)} = 8.893$, p = .003, $\eta_p^2 = 0.053$.

In contrast to hypothesis 5, as summarized in the fifth line of Table 2, there was no significant effect of media indicating that the VR groups (*M* = 1.28, *SD* = 0.87) did not differ significantly from the video groups (*M* = 1.27, S*D* = 0.86) on the transfer test $F_{(1,160)} = .009$, $p = .923$, and there was no significant interaction, $F_{(1,160)} = 2.738$, $p = .100$, $\eta_p^2 = 0.017$.

This pattern results concerning media effects replicates Experiment 1: learning in IVR increases enjoyment and presence but not learning. However, there was also an interaction between media and method indicating that enactment improved procedural knowledge more when learning in IVR than through a video.

Regarding hypothesis 6, Experiment 2 also investigated whether the generative learning strategy of enactment could improve performance on learning outcome tests, particularly for students learning in IVR. In support of this idea, there was a main effect for method showing that the groups with enactment (*M* = 8.06, *SD* = 4.87) scored significantly higher on procedural knowledge than the no enactment groups (*M* = 6.69, *SD* = 4.71), $F_{(1,160)} = 4.386$, p = .038, $\eta_p^2 = 0.027$. Since there was an interaction between media and method, $F_{(1,160)} = 8.893$, $p = .003$, $\eta_p^2 = 0.053$, the differences within each media condition were investigated independently. The results indicated that there was a significant difference showing that the enactment group (M = 8.76, SD = 4.62) scored significantly higher than the no enactment group (*M* = 5.27, *SD* = 4.48), $F_{(1,80)} = 13.501$, $p < .001$, $\eta_p^2 = 0.144$ when using IVR. However, the difference between the enactment (*M* = 7.40, *SD* = 5.07) and no enactment (*M* = 8.21, *SD* = 4.51) groups was not statistically significant in the video condition $F_{(1,79)} = .333$, $p = .566$, $\eta_p^2 = 0.004$. This is a major new finding in this experiment, and indicates that enactment can improve procedural knowledge, only when learning through IVR.

Related to the outcome of transfer, there was a main effect for method showing that the groups with enactment (*M* = 1.42, *SD* = 0.82) scored significantly higher than the no enactment groups (*M* = 1.12, *SD* = 0.89), $F_{(1,160)} = 4.686$, $p = .032$, $\eta_p^2 = 0.028$. Although the interaction between media and method was not significant $F_{(1,160)} = 2.738$, $p = .100$ $\eta_p^2 = 0.017$, when investigating the differences within each media condition independently, the results showed that there was a significant difference indicating that the enactment (*M* = 1.54, *SD* = 0.78) group scored significantly higher than the no enactment group (*M* = 1.02, *SD* = 0.90), when using IVR, $F_{(1,80)} = 7.676$, $p = .007$, $\eta_p^2 = 0.088$. However, the difference between the enactment (*M* = 1.30, *SD* = 0.86), and the no enactment (*M* = 1.23, *SD* = 0.87) groups did not reach statistical significance in the video condition, $F_{(1,79)} = .102$, $p = .750$, $\eta_p^2 = 0.001$. This strengthens the major new finding in Experiment 2, indicating that enactment improves transfer, when learning through IVR.

As expected, enactment did not affect performance in the declarative knowledge posttest, with enactment groups (*M* = 9.94, *SD* = 3.36) not scoring significantly differ than non-enactment groups (*M* = 10.21, *SD* = 4.81), $F_{(1,160)} = .002$, $p = .964$, $\eta_p^2 = 0.000$, and there was no significant interaction, $F_{(1,160)} = .365$, $p = .547$, $\eta_p^2 = 0.002$.

Overall, the results of Experiment 2 support hypothesis 6.

Discussion of Experiment 2

Experiment 2 investigated the relationship between instructional media (IVR vs. video) and method (enactment vs. no enactment). Consistent with Experiment 1, students who learned through IVR had significantly higher ratings of perceived enjoyment and presence following the lesson compared to the students who learned through the video. However, unlike Experiment 1, there was a significant main effect for media for the outcome of declarative knowledge. That is, students who learned in the video condition scored significantly better than those who learned in the IVR condition regardless of enactment. The results are inconsistent with the results from Experiment 1 where there was no significant difference on the declarative knowledge test between the IVR and video groups. Although the effect size was small $\eta_p^2 = 0.028$, the results support previous research which suggests that IVR is not optimal for declarative knowledge acquisition unless the lesson is implemented in a larger educational context. This could include providing students with pre-training material before the IVR lesson (Meyer et al., 2019), or by segmenting a IVR lesson, and having them summarize the material after each segment in order to encourage reflection and self-regulated learning (Parong & Mayer, 2018).

Adding the generative learning strategy of enactment to a lesson benefitted students' procedural knowledge and transfer in the IVR condition but did not have a significant effect on student's perceived enjoyment, presence, or declarative knowledge. This supports previous claims that it is specifically important to design IVR lessons within a larger educational context that includes an appropriately designed VR lesson, and a follow up activity such as a generative learning strategy (Meyer et al., 2019; Parong & Mayer, 2018). This is the case because IVR lessons can be very engaging, meaning that students are cognitively active when learning through the media (Makransky et al., 2019b). However, appropriate instructional design for VR lessons is essential to ensure that students' engagement leads to appropriate generative processing including the selection of relevant rather than irrelevant information, and its organization and integration with long term memory in a way that does not cognitively overload students. Even though the enactment procedure was relatively easy to perform, it helped the IVR group it promoted deep learning as measured by transfer test performance.

Furthermore, the higher level of presence afforded by the IVR lesson was closely related to the enactment exercise as the props where designed to mirror the virtual environment. Therefore, the students who felt like they experienced being in the lab in the IVR lesson could have had an advantage in recognizing and remembering their actions conducted with the props compared to the video condition where students saw the content without directly experiencing or interacting with it. Alternatively, it might be more difficult for students to imagine themselves performing the task after watching the video as they were not active in the learning experience, explaining why no significant effect of enactment was found in the video condition. A complementary explanation could be that enactment had a compensatory effect for the IVR lesson where students had suboptimal learning before given the opportunity to reflect over in the content through the generative leaning strategy of enactment.

General Discussion

Empirical Contributions

The primary pattern of results across two experiments is that students like learning in IVR and feel a greater sense of presence, but they do not learn better than from conventional media such as video. Experiment 2 added evidence that generative learning strategies such as enactment can be effective in improving learning, particularly in IVR. Finally, we found that there was an interaction between media and method for the outcome of procedural knowledge, suggesting that some instructional methods might be specifically relevant when learning through IVR.

Theoretical Implications

Consistent with most previous research on the use of IVR to acquire declarative knowledge, the results in this study demonstrated that the video was just as good (Experiment 1) or slightly better (Experiment 2) than the IVR simulation. The CTML describes how learning in multimedia can result in essential processing overload from the complexity of the material, as well as extraneous processing overload from the presentation of the material and that this load is additive (Makransky, Terkildsen & Mayer, 2019c; Sweller, Ayres, Kalyuga, 2011). The fact that the video condition outperformed the IVR condition in Experiment 2 but not Experiment 1 could be a consequence of students from the general high school population in Experiment 2 experiencing the content as being more difficult than those in the honors science camp who participated in Experiment 1. This is the case because extraneous processing can be harmful to the process of learning because working memory load is increased when the content is not directly related to the to-be-learned material, specifically in learning settings where students experience the material to be difficult (Mayer, 2014).

There were no main effect differences between the IVR and video conditions on procedural knowledge or transfer. IVR provides the learner with added interactivity and agency which should be beneficial to learning because they can actively take control of the pace, thereby engaging in higher generative processing (Makransky & Petersen, 2019). Previous studies (Alhalabi, 2016; Webster, 2016) have found students to learn more in immersion VR, partially due to the level of interactivity and control. For instance, Webster (2016) concludes that the interactivity in IVR, activated learners' senses which helped aid more learning. However, more control and agency can also have negative consequences for learning due to cognitive load that results from a more complicated learning environment or when the control mechanisms are not familiar (Makransky et al., 2019b; Sawyer et al., 2017). In the simulation used in this study the learner had the freedom to look in a 360-degree view in the laboratory environment, which could have added to the student's cognitive load as they also had to pay attention to the pedagogical agent, read relevant information, and perform tasks in the virtual environment. Therefore, some students might divert their attention from important material, essentially leading to students being unable to properly process the to-be-learned content. In the video condition students were taken through an optimal run of the simulation thereby only experiencing relevant information, which could have led to less generative processing which is beneficial to learning, but also less extraneous processing which is detrimental to learning.

Adding enactment as an instructional method to IVR led to significantly higher procedural knowledge and transfer scores. This provides evidence for generative learning theory in that enacting the material immediately after the lesson, prompted the learners to select, organize, and integrate the information from the lesson into their existing knowledge structures. However, enactment did not significantly benefit the video groups. A major finding in this study is therefore that generative learning strategies can successfully be applied to learning in IVR environments, and that they may be even more relevant than when learning with traditional media.

The current study further demonstrates that students feel more present and report higher enjoyment when learning through IVR. IVR thereby has the potential to trigger situational interest by making the learned subjects more interesting and relevant because students can experience working with the material in a realistic scenario. Although motivational factors are not always enough to enhance learning (Dewey, 2004), they can develop into later phases involving individual interest development which have been found to promote positive long-term educational outcomes (Renninger & Hidi, 2016).

Practical Implications

The results show that using the generative learning strategy of enactment is specifically beneficial for increasing procedural knowledge and transfer in IVR. Researchers and practitioners should therefore be aware of the value of using follow-up activities such as enactment after an IVR lesson. For example, a practical way of using enactment could be to include it as part of the regular classroom lesson, where students first engage in some form of pre-training (Meyer et al., 2019), then use an IVR lesson followed by enactment drills, where students can reflect and draw upon their new knowledge. This could be practical within classroom settings, since it is our experience that it is optimal to have a small group of students engage in the IVR lesson while others work on other relevant activities.

An important practical consideration is that IVR software and hardware is still costly and impractical in many educational settings. Although the results from this study suggest that teachers could get the same learning outcomes from projecting a video of a simulation to all participants, it is important to also consider the motivational outcomes in favor of IVR systems, which could cause students to persist longer in their learning activities. Since IVR is a popular technology among students, as indicated by the increase of students' perceived enjoyment, using it in combination with a generative learning strategy may be beneficial in fostering student motivation and still sustain learning outcomes when compared to other more traditional media.

The results also support previous research suggesting that it may not be optimal to present all learning content in IVR (e.g., Meyer et al., 2019), and that the ultimate value of IVR as a learning platform may be to combine it within more traditional media. Practitioners should be thoughtful in choosing which parts of a lesson to present in IVR. For instance, the results support previous research suggesting that declarative knowledge can be acquired more effectively in more traditional media (e.g., Makransky et al., in press; Parong & Mayer, 2018).

Simulation sickness is a practical issue that is often referred to in research related to IVR (e.g., Polcar & Horejsi, 2015; So, Lo & Ho, 2001). In this study, simulation sickness was limited by using a commercial product that had been thoroughly designed and tested, and by ensuring that students could stop the session if they felt uncomfortable. Simulation sickness did not cause any of the students to stop the session; however, a few students did mention that the HMD was not comfortable in post-intervention interviews. Practitioners should consider these factors in implementing IVR based lessons in applied settings.

Several factors could be optimized in the current IVR simulation that may potentially lead to better learning outcomes. There are promising innovative solutions to introduce sensations of touch (Whitmire, Benko, Holz, Ofek, & Sinclair, 2018), weight perception (Rietzler, Geiselhart, Gugenheimer & Rukzio, 2018), and force feedback (Choi & Follmer, 2016) in IVR which could potentially have a positive effect on procedural knowledge and skill acquisition. However, in this study we were limited to using a commercial product that could be used in a realistic high school science lesson that used a click and touch system rather than more advanced interaction. This allowed us to test a system that is currently being used in educational settings with the advantages of low simulation sickness; however, the fact that this system is unintuitive and unfamiliar to most students could have resulted in added extraneous load that negatively impacted learning.

Furthermore, since instructional design is important in IVR (e.g., Makransky et al., 2019b) several features could be adapted in the IVR simulation used in this study to optimize procedural knowledge acquisition and transfer. One option would be to introduce generative learning strategies within the IVR simulation to prompt metacognition and reflection. This may be specifically relevant for the current simulation as some students might have followed through the guided fixed linear sequence of problem-solving activities without reflecting about the procedures or process.

The teacher's role is an aspect that cannot be ignored, as teachers are in many ways definitive for students' learning processes. As presented in the article, the results emphasize that IVR should not be used as a standalone tool, but instead used thoughtfully as part of the classroom teaching. Therefore, the biggest challenge is arguably to help teachers accept the technology, and understand the particular affordances and ways in which it can support learning.

Future Directions and Limitations

A potential limitation in this study is that the correlation between the ratings in Experiment 1 was 0.74. The correlation increased to an acceptable level of 0.90 in Experiment 2. An explanation for this difference could be that the honors students in Experiment 1 wrote significantly more text than the general sample of high school students in Experiment 2 making it more difficult to assign scores. Future research should investigate the generalizability of the results using different forms of assessment, in different samples, and with different IVR simulation.

The enactment exercise in Experiment 2 resulted in a longer intervention and more time-on-task than the control condition. Although this only resulted in increased procedural knowledge and transfer outcomes in the IVR group, future research should investigate if this time on task could be used more effectively. For instance, research could investigate if enactment following IVR is more effective than using other GLS such as summarizing, teaching, or simply having students re-study the material.

Another potential limitation in this study is that students interacted with features in the virtual environment through a touchpad on the side of the head because we prioritized using a commercially available IVR simulation which is currently used in educational contexts. More research is needed to investigate if IVR simulations that afford higher interaction fidelity through gloves or hand trackers would result in different learning outcomes. A further potential limitation in this study was that we did not collect log data about student's behavior while learning. Future research could use log files, or collect psychophysiological data to further understand the processes underlying learning in IVR and video.

Several of the points brought up in this discussion entail an understanding of IVR as a medium, which may augment or limit the importance of instructional methods differently than a video. More research is needed to determine if evidence based instructional methods generalize to learning in IVR, as well as how other generative learning strategies can be used to help students promote deeper learning in IVR. Further research should also compare the value of using different generative learning strategies using different learning outcomes. More longitudinal studies are also needed to investigate the learning effectiveness of IVR as higher levels of perceived enjoyment and presence could lead to long-term effects on learning outcomes. Finally, the results in this study are interpreted based on known motivational and cognitive theoretical frameworks, nonetheless, more research is needed to investigate the process of learning in IVR in order to further develop these theoretical frameworks. This could be achieved by combining psychophysiological measures with log files to gain a better understanding of how factors such as enjoyment, presence, and cognitive load lead to learning.

References

Abrahamson, D., Sánchez-García, R., & Abrahamson, D. (2016). Learning is moving in new ways : The ecological dynamics of mathematics education. *Journal of the Learning Sciences*, *25*(2), 203–239. https://doi.org/10.1080/10508406.2016.1143370

Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of educational psychology*, *104*(1), 235-249.

Aggarwal, R., Tully, A., Grantcharov, T., Larsen, C. R., Miskry, T., Farthing, A., & Darzi, A. (2006). Virtual reality simulation training can improve technical skills during laparoscopic salpingectomy for ectopic pregnancy. *BJOG: An International Journal of Obstetrics & Gynaecology*, *113*(12), 1382-1387.

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

Andreasen, N. K., Baceviciute, S., Pande, P., & Makransky, G. (2019, March). Virtual Reality Instruction Followed by Enactment Can Increase Procedural Knowledge in a Science Lesson. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 840-841). IEEE.

Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do.* New York: Norton & Company.

Barfield, W., Zeltzer, D., Sheridan, T.B., Slater, M. (1995). Presence and performance within virtual environments. In W. Barfield & T. A. Furness (Eds.), *Virtual environments and advanced interface design* (pp. 473–541). New York: Oxford University Press.

Baceviciute, S., Mottelson, A., Terkildsen, T., & Makransky, G. (2020). Investigating Representation of Text and Audio in Educational VR using Learning Outcomes and EEG. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems

(CHI '20). ACM, New York, NY, USA. http://dx.doi.org/10.1145/3313831.3376872

Bonde, M. T., Makransky, G., Wandall, J., Larsen, M. V., Morsing, M., Jarmer, H., & Sommer, M. O. A. (2014). Improving biotech education through gamified laboratory simulations. *Nature Biotechnology, 32*(7), 694–697. https://doi.org/10.1038/nbt.2955

Brom, C., Dobrovolný, V., Dechterenko, F., Stárková, T., & Bromová, E. (2019). It's Better to Enjoy Learning than Playing: Motivational Effects of an Educational Live Action Role-Playing Game. *Frontline Learning Research, 7*(3), 64-90.

Buttussi, F., & Chittaro, L. (2018). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE Transactions on Visualization and Computer Graphics*, *24*(2), 1063–1076. https://doi.org/10.1109/TVCG.2017.2653117

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380–400. https://doi.org/10.1037/a0031084

Chittaro, L., & Sioni, R. (2015). Serious games for emergency preparedness: Evaluation of an interactive vs. a non-interactive simulation of a terror attack. *Computers in Human behavior*, 50, 508-519.

Choi, I & Follmer, S. 2016. Wolverine: A wearable haptic interface for grasping in VR. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 117-119). New York: ACM.

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Dewey, J. (1913). *Interest and effort in education*. Cambridge, MA: Houghton Mifflin.

Dewey, J. (2004). *Democracy and education*. Mineola, N.Y: Dover Publications.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.

Fiorella, L., & Mayer, R. E. (2012). Paper-based aids for learning with a computer- based game. *Journal of Educational Psychology*, *104*(4), 1074–1082. https://doi.org/10.1037/a0028088

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding.* New York: Cambridge University Press.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Freeman, A., Becker, S. A., & Cummins, M. (2017). *NMC/CoSN horizon report: 2017 K*. Austin, Texas: The New Media Consortium.

Freina, L., and Ott, M. (2015). “A literature review on immersive virtual reality in education: state of the art and perspectives,” in *Proceedings of eLearning and Software for Education (eLSE), 2015 April 23–24*. Bucharest.

Fujimura, N. (2001). Facilitating children’s proportional reasoning: A model of reasoning processes and effects of intervention on strategy change. *Journal of Educational Psychology*, *93*(3), 589–603. https://doi.org/10.1037/0022-0663.93.3.589

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology*, *45*(6), 1097–1114. https://doi.org/10.1111/bjet.12188

Glenberg, A. M., Goldberg, A. B., & Zhu, X. (2011). Improving early reading comprehension using embodied CAI. *Instructional Science*, *39*(1), 27–39. https://doi.org/10.1007/s11251-009-9096-7

Howard, M. (2019). Virtual Reality Interventions for Personal Development: A Meta-Analysis of Hardware and Software. *Human–Computer Interaction*, *34*(3), 205-239.

IJsselsteijn, W. A., de Ridder, H., Freeman, J., & Avons, S. E. (2000, January). Presence: Concept, determinants and measurement. Proceedings of the SPIE, Human Vision and

Electronic Imaging V, San Jose, CA.

Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive vieweing for leanring anatomy in three-dimentional virtual reality environment. *Computers & Education*, 106, 150 - 165.

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.

John, N. W., Pop, S. R., Day, T. W., Ritsos, P. D., & Headleand, C. J. (2018). The Implementation and Validation of a Virtual Environment for Training Powered Wheelchair Manoeuvres. *IEEE Transactions on Visualization and Computer Graphics*, *24*(5), 1867–1878. https://doi.org/10.1109/TVCG.2017.2700273

Johnson-Glenberg, M. C. (2019). The Necessary Nine: Design Principles for Embodied VR and Active Stem Education. In *Learning in a Digital World* (pp. 83-112). Springer, Singapore.

Kirschner, P., Sweller, J., Clark, R. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist.* 41, 75–86.

Labster. (2019). Labster - Polymerase Chain Reaction lab. Retrieved August 12th 2019 from: https://www.labster.com/simulations/polymerase-chain-reaction/

Leder, J., Horlitz, T., Puschmann, P., Wittstock, V., & Schütz, A. (2019). Comparing immersive virtual reality and powerpoint as methods for delivering safety training : Impacts on risk perception , learning , and decision making. *Safety Science*, *111*(February 2018), 271–286. https://doi.org/10.1016/j.ssci.2018.07.021

Lee, K. M. (2004). Presence, explicated. *Communication Theory*, *14*(1), 27e50.

Lele, A. (2013). Virtual reality and its military utility. *Journal of Ambient Intelligence and Humanized Computing*, *4*(1), 17–26. https://doi.org/10.1007/s12652- 011-0052-4

Li, C., Liang, W., Quigley, C., Zhao, Y., & Yu, L. F. (2017). Earthquake Safety Training through Virtual Drills. *IEEE Transactions on Visualization and Computer Graphics*, *23*(4), 1388–1397. https://doi.org/10.1109/TVCG.2017.2656958

Makransky, G., Borre-Gude S., & Mayer, R. E. (in press). Motivational and Cognitive Benefits of Training in Immersive Virtual Reality Based on Multiple Assessments. *Journal of Computer Assisted Learning*. DOI: 10.1111/jcal.12375

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, *66*(5), 1–24. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, *72*, 276–285. https://doi.org/10.1016/j.chb.2017.02.066

Makransky, G., & Petersen, G. B. (2019). Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers and Education*, *134*, 15–30. https://doi.org/10.1016/j.compedu.2019.02.002

Makransky, G., Mayer, R. E., Nørremølle, A., Lopez Cordoba, A., Wandall, J., Bonde, M. (2019b). Investigating the feasibility of using assessment and explanatory feedback in desktop virtual reality simulations. *Educational Technology Research and Development.* https://doi.org/10.1007/s11423-019-09690-3.

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

Makransky, G., Terkildsen, T. S., & Mayer. R. E. (2019c). Role of subjective and objective measures of cognitive processing during learning in explaining the spatial contiguity effect. *Learning and Instruction, 61,* 23-34. https://doi.org/10.1016/j.learninstruc.2018.12.001

Makransky, G., Wismer, P. Mayer, R. E. (2018). A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of*

Computer Assisted Learning, *35*, 349-358. https://doi.org/10.1111/jcal.12335

Mayer, R. E. (2004). Should There Be a Three-Strikes Rule Against Pure Discovery Learning? The case for guided methods of instruction. *American Psychologist.* 59, 14–19.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed). New York: Cambridge University Press.

Mayer, R. E. (2014a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed; pp. 43-71). New York: Cambridge University Press.

Mayer, R. E. (2014b). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed; pp. 345-368). New York: Cambridge University Press.

Mayer, R. E. (Ed.). (2014c). *The Cambridge handbook of multimedia learning* (2nd ed.). New York: Cambridge University Press.

Meyer O. A., Omdahl M. K. & Makransky G., (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*. doi: https://doi.org/10.1016/j.compedu.2019.103603.

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten- year review of empirical research (1999-2009). *Computers and Education*, *56*(3), 769–780. https://doi.org/10.1016/j.compedu.2010.10.020

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, *94*(3), 598–610. https://doi.org/10.1037/0022-0663.94.3.598

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110, 785-797. https://doi.org/10.1037/edu0000241

Pekrun, R. (2006). The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology*, *36*(1), 36-48.

Pekrun, R., & Stephens, E. J. (2010). Achievement emotions: A control-value approach. Social and Personality Psychology Compass, 4(4), 238–255. https://doi.org/10.1111/j.1751-9004.2010.00259.

Pilegard, C., & Mayer, R. E. (2016). Improving academic learning from computer- based narrative games. *Contemporary Educational Psychology*, *44–45*, 12–20. https://doi.org/10.1016/j.cedpsych.2015.12.002

Polcar, J., & Horejsi, P. (2015). Knowledge acquisition and cyber sickness: a comparison of VR devices in virtual tours. *MM Science Journal*, 613–616. https://doi.org/10.17973/MMSJ.2015_06_201516.

Renninger, K. A., & Hidi, S. (2016). *The power of interest for motivation and engagement.* New York: Routledge.

Richards, D., & Taylor, M. (2015). A comparison of learning gains when using a 2D simulation tool versus a 3D virtual world: An experiment to find the right representation involving the Marginal Value Theorem. *Computers and Education*, *86*, 157–171. https://doi.org/10.1016/j.compedu.2015.03.009

Riva, G. (2008). From virtual to real body: virtual reality as embodied technology. *Journal of Cyber Therapy Rehabilitation*, *1*, 7-22.

Ritterfeld, U., Shen, C., Wang, H., Nocera, L., & Wong, W. L. (2009). Multimodality and interactivity: Connecting properties of serious games with educational outcomes. *Cyberpsychology & Behavior*, 12(6), 691-697.

Rothbaum, B. O., Hodges, L. F., Ready, D., Graap, K., & Alarcon, R. D. (2001).Virtual reality exposure therapy for Vietnam veterans with posttraumatic stress disorder. *The Journal of Clinical Psychiatry*, *62(8), 617-622.*

Sacks, R., Perlman, A., & Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics*, *31*(9), 1005-1017.

Sawyer, R., Smith, A., Rowe, J., Azevedo, R., & Lester, J. (2017). Is more agency better? The impact of student agency on game-based learning International Conference on Artificial Intelligence in Education (pp. 335-346): Springer.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183-212). Hillsdale, NJ: Erlbaum.

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 197-223*)*. New York: Routledge.

Schrader, C., & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human behavior, 28*(2), 648-658.

Shin, D. H. (2017). The role of affordance in the experience of virtual reality learning: Technological and affective affordances in virtual reality. *Telematics and Informatics*, *34*(8), 1826-1836.

Slobounov, S. M., Ray, W., Johnson, B., Slobounov, E., & Newell, K. M. (2015). Modulation of cortical activity in 2D versus 3D virtual reality environments: An EEG study. *International Journal of Psychophysiology*, *95*(3), 254–260. https://doi.org/10.1016/j.ijpsycho.2014.11.003

So, R. H., Lo, W. T., & Ho, A. T. (2001). Effects of navigation speed on motion sickness caused by an immersive virtual environment. *Human factors*, *43*(3), 452-461.

Song, H. S., Pusic, M., Nick, M. W., Sarpel, U., Plass, J. L., & Kalet, A. L. (2014). The cognitive impact of interactive design features for learning complex materials in medical education. *Computers & Education*, 71, 198-205.

Steinemann, S. T., Iten, G. H., Opwis, K., Forde, S. F., Frasseck, L., & Mekler, E. D. (2017). Interactive Narratives Affecting Social Change. *Journal of Media Psychology*, 29, 54-66

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York: Springer

Terkildsen, T. S., Makransky, G. (2019). Measuring Presence in Video Games: An

Investigation of the Potential use of Physiological Measures as Indicators of Presence. *International Journal of Human Computer Studies.* https://doi.org/10.1016/j.ijhcs.2019.02.006

Thisgaard, M., & Makransky, G. (2017). Virtual learning simulations in high school: Effects on cognitive and non-cognitive outcomes and implications on the development of STEM academic and career choice. *Frontiers in Psychology*, *8*(MAY), 1–13. https://doi.org/10.3389/fpsyg.2017.00805

Tokel, S. T., & İsler, V. (2013). Acceptance of virtual worlds as learning space. *Innovations in Education and Teaching International*, *3297*(January 2014), 1–11. https://doi.org/10.1080/14703297.2013.820139

van der Heijden, H. 2004. "User Acceptance of Hedonic Information Systems," *MIS Quarterly*, *28:4*, 695-704.

van Ginkel, S., Gulikers, J., Biemans, H., Noroozi, O., Roozen, M., Bos, T., … Mulder, M. (2019). Computers & Education Fostering oral presentation competence through a virtual reality- based task for delivering feedback. *Computers & Education*, *134*, 78–97. https://doi.org/10.1016/j.compedu.2019.02.006

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, *34*(3), 229-243.

Wardrip-Fruin, N., Mateas, M., Dow, S., & Sali, S. (2009, September). Agency reconsidered. In Breaking New Ground: Innovation in Games, Play, Practice and Theory: *Proceedings of the 2009 Digital Games Research Association Conference*,1-9. London, UK

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

Wigfield, A., Tonks, S., & Klauda, S. (2016). Expectancy-value theory. In *Handbook of motivation in school(pp. 55-74)*. New York, NY: Routledge.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Winne, P., Azevedo, R. (2014). Metacognition. In: Sawyer, K. (ed.) Cambridge Handbook of the Learning Sciences. pp. 63–87. Cambridge University Press, Cambridge, MA.

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist*, 11(2), 3–11. https://doi.org/10.1007/978-981-4021-93-7_1

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376.

Table 1. *Mean and Standard Deviation on the Four Dependent Measures for Two Groups--Experiment 1*

Measure	**Group**				**Sig.**	**Cohens d**
	IVR		**Video**			
	M	**SD**	**M**	**SD**		
1. Perceived enjoyment	4.22	(.56)	3.67	(0.84)	*p*<.001	d = 0.79
2. Presence	3.64	(.48)	2.92	(.66)	*p*=.001	d = 1.26
3. Declarative knowledge	14.31	(2.93)	14.35	(2.59)	*p*=.925	d = 0.01
4. Procedural knowledge	10.47	(2.79)	10.07	(2.58)	*p*=.402	d = 0.15

Table 2: *Mean and Standard Deviation (in Parentheses) on Five Dependent Variables for Four Groups--Experiment 2.*

	Groups						
	IVR		**Video**		**Media**	**Method**	**Interaction**
Enactment	**With**	**Without**	**With**	**Without**	**p-value**	**p-value**	**p-value**
1. Enjoyment	4.00 (0.73)	3.92 (0.95)	3.68 (0.85)	3.58 (0.91)	**.047**	.316	.961
2. Presence	3.11 (0.64)	3.03 (0.86)	2.51 (0.77)	2.33 (0.71)	.000	.313	.684
3. Declarative Knowledge	9.63 (3.23)	9.57 (4.61)	10.23 (3.49)	10.89 (4.98)	**.034**	.964	.547
4. Procedural Knowledge	8.76 (4.62)	5.27 (4.48)	7.40 (5.07)	8.22 (4.51)	.159	**.038**	**.003**
5. Transfer	1.54 (0.78)	1.02 (0.90)	1.30 (0.86)	1.23 (0.87)	.923	**.032**	.100

Figure 1: Screenshots of the CSI simulation

Figure 2: Picture of students in learning through VR (left panel) and video (right panel in Experiment 1.

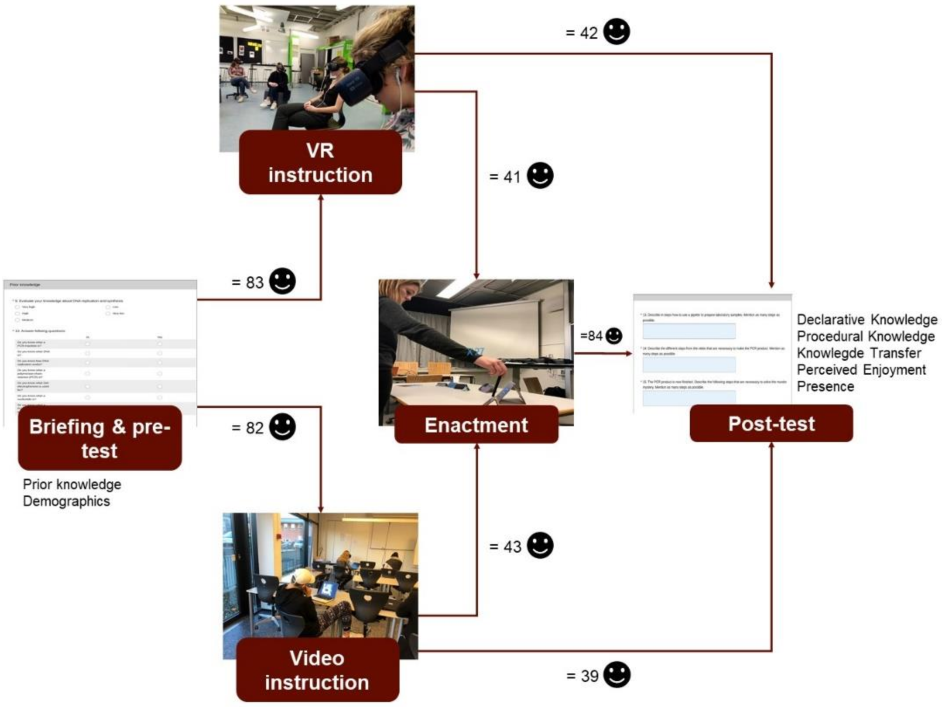

Figure 3: Experimental procedure in Experiment 2

Figure 4: Picture of students in learning through VR (left panel) and video (right panel in Experiment 2.

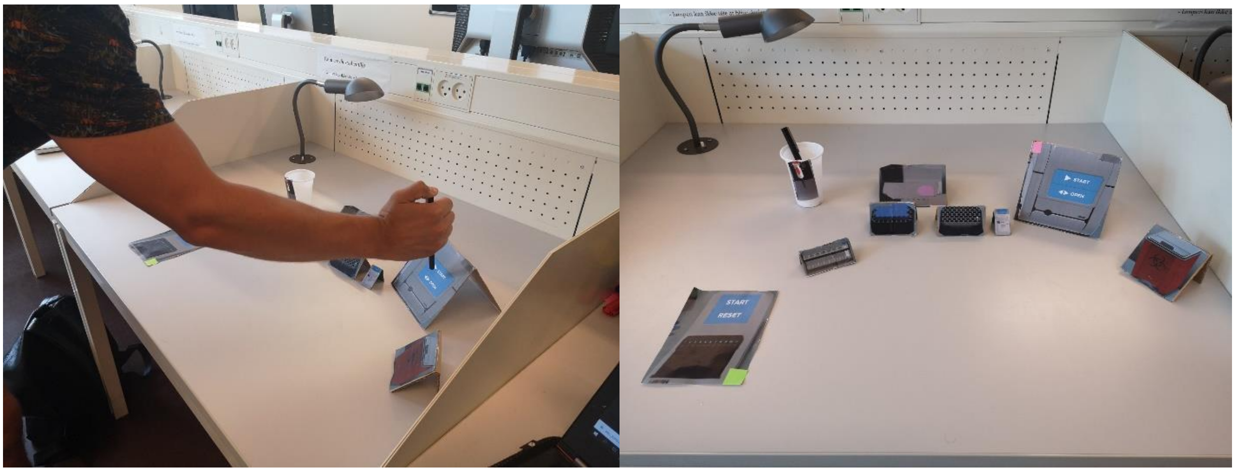

Figure 5: The left panel shows a photo of a student enacting. The right panel shows the enactment props.