Received: 30 March 2020 | Revised: 1 June 2020 | Accepted: 12 July 2020

DOI: 10.1111/jcal.12482

ARTICLE

Journal of Computer Assisted Learning WILEY

Cognitive and affective processes for learning science in immersive virtual reality

Jocelyn Parong[1,2] | Richard E. Mayer[1]

[1]Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, California, USA

[2]Department of Psychology, University of Wisconsin–Madison, Madison, Wisconsin, USA

Correspondence
Jocelyn Parong, Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706.
Email: jparong@wisc.edu

Funding information
Office of Naval Research, Grant/Award Number: N000141612046

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12482.

Abstract

As immersive virtual reality (IVR) systems proliferate in classrooms, it is important to understand how they affect learning outcomes and the underlying affective and cognitive processes that may cause these outcomes. Proponents argue that IVR could improve learning by increasing positive affective and cognitive processing, thereby supporting improved performance on tests of learning outcome, whereas opponents of IVR contend that it could hurt learning by increasing distraction, thereby disrupting cognitive learning processes and leading to poorer learning outcomes. In a media comparison study, students viewed a biology lesson either as an interactive animated journey in IVR or as a slideshow on a desktop monitor. Those who viewed the IVR lesson performed significantly worse on transfer tests, reported higher emotional arousal, reported more extraneous cognitive load and showed less engagement based on EEG measures than those who viewed the slideshow lesson, with or without practice questions added to the lessons. Mediational analyses showed that the lower retention scores for the IVR lesson were related to an increase in self-reported extraneous cognitive load and emotional arousal. These results support the notion that immersive environments create high affective and cognitive distraction, which leads to poorer learning outcomes than desktop environments.

KEYWORDS

affective processing, cognitive processing, multimedia, science learning, virtual reality

1 | INTRODUCTION

1.1 | Objective and rationale

Immersive virtual reality (IVR) is a technology that allows the user to experience physical and/or behavioural simulations (e.g., interacting with objects and environments; LaValle, 2019), thereby giving students access to experiences that otherwise are difficult or impossible to obtain (Bailenson, 2018). IVR consoles, such as the Oculus Rift or HTC Vive, have become more affordable and accessible, and they have gained popularity in classrooms among both students and teachers. However, a recent review of the IVR research reported that very few studies have empirically examined the effects of the technology on learning outcomes, but instead many focused on usability of the technology (Radianti, Majchrzak, Fromm, & Wohlgenannt, 2020). Thus, it is important to examine whether IVR can foster instruction that is more effective than with more traditional instructional media, such as a PowerPoint slideshow presentation.

On one hand, it has been proposed that using IVR to deliver academic lessons may promote affective processes that pique student's interest in learning (Bailenson, 2018; Bailenson & Blascovich, 2011). Indeed, many studies and consumer surveys report both high student and teacher support for using IVR in classrooms (Hew & Cheung, 2010; Makransky, Terkildsen, & Mayer, 2017; Parong & Mayer, 2018); for example, 83% of teachers believe IVR will improve learning outcomes (Samsung Business USA, 2016).

On the other hand, IVR may also have negative effects on cognitive or affective processing during the lesson, which, in turn, hurt learning. For example, certain features of the IVR environment may be distracting to learning processes, such as disrupting cognitive processing during information acquisition or creating excessive emotional arousal that detracts from learning. If this is the case, it would also be important to determine whether learning outcomes from IVR could be improved. One example could be asking students to engage in a generative learning strategy (Fiorella & Mayer, 2015), such as answering practice questions during the lesson. The goals of this study were to examine these issues.

1.2 | The case for IVR: Could IVR help learning?

IVR offers some practical and theoretical affordances over other instructional media. Possibly the biggest practical affordance is the wide range of lessons that could be displayed in IVR. These virtual learning environments can simulate any location combined with text, narrators and/or pedagogical agents to facilitate learning. For example, learners can gain access to previously inaccessible environments, such as travelling to space or entering an animal cell, or practice dangerous or expensive tasks, such as a chemistry experiment or a surgical procedure (Ott & Freina, 2015). In addition, IVR may offer unique affordances. The immersiveness and interactivity related to using IVR may increase learning outcomes by increasing a learner's affective processes and boosting cognitive processes associated with high motivation, such as attention and focus. Concerning affective processes, IVR could increase the learner's feelings of presence during a lesson and heighten the learner's level of emotional arousal experienced during a lesson. Concerning cognitive processes, IVR could increase the amount of cognitive activity during learning.

First, the level of immersion in many IVR experiences is associated with a user's sense of presence. Immersion and presence are closely related but distinct concepts. Immersion is an objective property of the physical technology, in which some technologies (such as immersive virtual reality using a head-mounted display) are considered to involve higher levels of immersion than other technologies (such as a desktop computer or printed book). In contrast, presence refers to the subjective feeling of "being there" (Nilsson, Nordhal, & Serafin, 2016). Learning environments with higher objective immersive qualities, such as higher resolution head tracking, faster update rates, stereoscopic vision (rather than monoscopic vision), greater degrees of freedom of head rotation and tracking, higher image and sound quality, more external visual occlusion and/or a larger field of view compared to environments with less immersive qualities, tend to elicit greater psychological presence. A meta-analysis reported that across different levels of immersive technologies, immersion had a medium-sized effect on presence (Cummings & Bailenson, 2016). Previous research has shown that presence is associated with learning outcomes in some cases (e.g., Makransky & Lilleholt, 2018; Schrader & Bastiaens, 2012). For example, Schrader and Bastiaens (2012) presented students with a physics lesson in either a 3D or 2D simulation. Learners in the 3D lesson reported more presence, and the amount of presence felt was positively associated with the performance on transfer tests. This sense of presence in IVR may increase motivation to be engaged and attentive in the simulation. In turn, learners would be more motivated to persist in a lesson and, in turn, increase outcomes from the goals of the lesson.

In addition, IVR lessons may induce positive emotions; for example, students generally report liking IVR lessons (e.g., Parong & Mayer, 2018). Extant research on emotions and memory has shown that positive emotions facilitate encoding during learning and help retrieval of information (see Storbeck & Clore, 2008; Tyng, Amin, Saad, & Malik, 2017 for reviews). Previous research has shown that the emotional design of a multimedia lesson can induce positive emotions in learners that facilitate comprehension and transfer. For example, in a lesson about immunization, using characters with round faces and warm colours induced positive emotions in learners, which, in turn, improved learning outcomes from the lesson compared to characters with square faces with no colour (Mayer & Estrella, 2014; Plass, Heidig, Hayward, Homer, & Um, 2014; Um, Plass, Hayward, & Homer, 2012). Taken together, using IVR may increase the learning outcomes by increasing motivation associated with feeling present and positive emotions in the immersive environment.

Supporting these ideas is some evidence from media comparison studies. Kozhevnikov et al. (2013) found that students who viewed a lesson on relative motion concepts in an immersive virtual environment using a head-mounted display did significantly better on transfer tests than students who learned from a desktop virtual environment presented on a computer screen. Similarly, Markowitz, Laha, Perone, Pea, and Bailenson (2018) found that students experienced knowledge gains and inquisitiveness about climate change after viewing an IVR lesson on climate change. Finally, Makransky, Borre-Gude, and Mayer (2019) also found that IVR was more effective in addressing problems in a real-world lab-setting than a text lesson.

1.3 | The case against IVR: Could IVR hurt learning?

In contrast, features of IVR may serve to detract from learning processes by causing cognitive or affective distraction. The cognitive theory of multimedia learning (CTML; Mayer, 2014), which is derived from cognitive load theory (Sweller, Ayres, & Kalyuga, 2011), describes three types of cognitive load a learner experiences during learning—extraneous load, which comes from cognitive processing that does not serve an instructional goal; essential load, which comes from processing relevant information; and generative load, which comes from more deeply processing the information. A learner's limited cognitive capacity can be allocated among these three cognitive processes, and effective learning occurs when cognitive resources are more focused on essential and generative processing, and less focused on extraneous processing (Fiorella & Mayer, 2015; Mayer, 2009; Mayer & Fiorella, 2014; Mayer & Pilegard, 2014).

Lessons in IVR may induce more extraneous cognitive processing than other media because they are more perceptually rich and present users with stimuli that are not relevant to the instructional goals, such as background music, animations, excessive perceptual detail or interesting but unnecessary interactions. The coherence principle from CTML states removing these extraneous stimuli improves learning outcomes (Mayer, 2009; Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Mayer & Fiorella, 2014). Extant research on the coherence principle shows its effectiveness across various domains and media, with a median effect size of d = .83 (Mayer & Fiorella, 2014). Given the limited capacity of working memory, this additional extraneous processing may, in turn, reduce cognitive resources available to devote to essential and generative processing, which would lead to poorer learning outcomes.

In addition, IVR could induce excessive affective processing, which could also be a source of extraneous cognitive processing. Although some positive emotions benefit learning as described earlier, more extreme positive emotions can be detrimental to learning. Based on Russell's (2003) description of core affect, emotions lie on two dimensions, pleasure and activation (i.e., arousal). High arousal events (involving both negative and positive pleasures) are often reported to be remembered more vividly than neutral events (e.g., Canli, Desmond, Zhao, & Gabrieli, 2002; Charles, Mather, & Carstensen, 2003; Ochsner, 2000), but the memories are also less accurate (Schmolck, Buffalo, & Squire, 2000). One study found that memory for pictures decreased as arousal increased; it was proposed that this was because emotional arousal disrupts encoding processes for binding features together in working memory for short-term retention (Mather et al., 2006).

Emotional arousal as a distractor is in line with CTML. Excessive emotional arousal may act as a source of extraneous cognitive processing, as it diverts cognitive resources to processing the emotion-evoking stimuli or the emotions themselves. Students may report feeling more high arousal emotions (e.g., excited or stressed vs. calm or bored) during an IVR lesson compared to other media, which may decrease their learning outcomes. In addition, high arousal emotions could cause task-irrelevant processing in which highly arousing emotions lead to the retrieval of irrelevant information that is not essential to learning (Plass & Kalyuga, 2019).

Previous research also supports the notion that IVR is distracting. Parong and Mayer (2018) found that viewing a cell biology lesson in IVR led to lower learning outcomes than viewing an equivalent lesson in a self-paced PowerPoint presentation. Other studies also have found similar results in comparing a science lesson in IVR to traditional media (Makransky et al., 2017; Moreno & Mayer, 2004).

1.4 | The case for improving IVR: Could IVR become more effective?

It may also be important to examine whether adding elements to existing IVR lessons could improve learning outcomes. Previous studies typically report students enjoying the use of IVR in classroom, so it may be worthwhile to determine how to improve the effectiveness of IVR lessons (Makransky et al., 2017; Moreno & Mayer, 2004; Parong & Mayer, 2018). Generative learning theory suggests that by increasing generative cognitive processing, for example, through learner-based activities, learning outcomes should be improved (Fiorella & Mayer, 2015). One example of a generative learning strategy is practice testing, which can be defined as low-stake or no-stakes prompts to recall or transfer previously learned material; some examples include practicing recall with flashcards or completing practice problems that may or may not be similar to problems on the final exam (Brown, Rodiger, & McDaniel, 2014; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Fiorella & Mayer, 2015; Miyatsu, Nguyen, & McDaniel, 2018). Previous research over the past century has shown robust effects of practice testing (e.g., Spitzer, 1939, Roediger & Karpicke, 2006; see Roediger & Butler, 2011 for a review). A secondary aim of this study is to determine whether the effects of practice testing can be extended to academic lessons in IVR.

1.5 | Current study

Figure 1 summarizes four components of learning with media that are relevant to the current study. First, in terms of instructional media, we compare learning from a lesson in IVR using an HTC Vive or from a PowerPoint slideshow on a desktop computer; and in terms of instructional methods, we compare learning with and without the generative activity of answering practice questions. A biology lesson on the human bloodstream was selected as it had previously been used in a research study (Parong & Mayer, 2018). Second, in terms of affective processes, we measure learners' self-reported ratings of presence, affective state, interest, enjoyment and motivation, as well as biometric measures of electrodermal activity (EDA), heart rate (HR), and heart rate variability (HRV). Third, in terms of cognitive processes, we measure learners' self-reported ratings of cognitive load as well as brain activity measures based on electroencephalograms (EEG). Fourth, in terms of learning outcomes, we measure post-test performance on retention and transfer.

The current study examines three research questions: (a) Is a biology lesson in IVR as effective as an equivalent lesson in a PowerPoint lesson on a desktop computer? (b) What are the cognitive and affective processes that can shape the learning outcomes from these media? (c) Does adding a generative learning strategy of practice testing during the lesson boost learning outcomes in either medium? This study aimed to replicate previous findings, that a biology lesson was more effective in promoting post-test performance when presented as a PowerPoint presentation on a desktop computer than as an interactive immersive simulation in IVR (Parong & Mayer, 2018), and to extend the findings to determine the underlying cognitive and affective processes associated with learning in each medium. In addition, the practice testing effect was examined in the form of oral questions asked throughout the lesson with both IVR and desktop media.

Based on CTML, the IVR and desktop lessons differ in the amount of extraneous cognitive load required to process the lesson; the IVR lesson includes extraneous animations, sounds, etc. that are omitted in the desktop lesson. However, the amount of essential and generative load should theoretically not differ between the two, as the instructional content, learning activities and feedback in the base

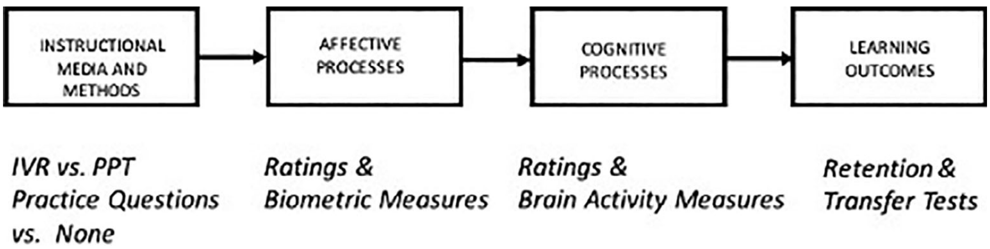

FIGURE 1 Four components in learning with media

lesson remains the same between the two. Therefore, we tested the distraction hypothesis, which proposes that IVR causes learners to use some of their limited cognitive resources on cognitive or affective processing that does not support deep learning of the material, as compared to more conventional media.

First, the distraction hypothesis predicts that learners will perform worse on retention and transfer tests after learning in IVR than learning the same material as a slideshow on a desktop computer (prediction 1). Second, we expect students, who learn in IVR, to report higher levels of emotional arousal (prediction 2a) and display higher levels of emotional arousal on biometric measures of HR and skin conductance (prediction 2b) as compared to students who learn with a slideshow. Third, we expect students, who learn in IVR, to report higher levels of extraneous cognitive processing (prediction 3a) and display less engagement based on brain activity measures (prediction 3b) as compared to students who learn with a slideshow. Whereas prediction 1 reflects a replication of the results of Parong and Mayer (2018), predictions 2a, 2b, 3a and 3c aim to extend the previous work by examining the affective and cognitive processes primed by learning in IVR.

Predictions 2a and 3a were assessed using self-report ratings, whereas predictions 2b and 3b were assessed using biometric measures. To address prediction 2b, we included measures of EDA, heart rate and HRV. EDA measures, including average skin conductance level, skin conductance responses per second and the average amount of change over time (peak-to-peak amplitude) have been validated as a measure of emotional arousal (e.g., Andreassi, 2007; Bradley & Lang, 2007; Lang, Greenwald, Bradley, & Hamm, 1993). HR and HRV (i.e., changes in heart rate) have also been associated with activity from the autonomic nervous system and regulated emotional responses (Appelhans & Luecken, 2006; Kim, Bang, & Kim, 2004). Lower HRV, specifically in the high-frequency band, indicates higher arousal and has been found to be correlated with stress, panic, anxiety and worry (Kim, Cheon, Bai, Lee, & Koo, 2018; Shaffer & Ginsberg, 2017). We predicted that the IVR groups would exhibit higher indications of physiological arousal according to EDA and HR measures than PowerPoint groups (prediction 2b).

To address the prediction in 3b, we included electroencephalography-based (EEG) measures of cognitive processing during the lesson. Previous research has shown that changes across the electrical activity can serve as direct and measurable indices of specific brain activities, such as attention, learning and memory (Basar, 1999). Advanced Brain Monitoring's B-Alert X10 mobile EEG system included proprietary measures of cognitive engagement and cognitive workload, which have been shown to be valid indexes of cognitive processing (Berka et al., 2004, 2007; Poythress et al., 2006). Similar to prediction 3a, we hypothesized those who viewed the IVR lesson would exhibit lower probabilities of being in a high cognitive engagement state and there would be fewer participants classified to have ideal cognitive workload (compared to boredom or cognitive overload) during learning than students who viewed the PowerPoint slideshow (prediction 2b).

A proposed cognitive-affective distraction model of IVR learning environments is shown in Figure 2a. This mediation model predicted that learners would be more emotionally and cognitively aroused in IVR than in a desktop lesson, which would cause poorer learning outcomes. Specifically, self-reported positive/high arousal emotions and HRV were used as measures of affective processing, and self-reported extraneous cognitive load and probability of high cognitive engagement were used as measures of cognitive processing, as shown in Figure 2b. As immersion increases (comparing desktop slideshow to IVR), extraneous cognitive processing and extraneous affective processing will increase, which will cause a decrease in learning (prediction 4).

A secondary aim of this study was to investigate whether a generative learning strategy, practice testing, could be extended to a novel medium in two academic domains. Based on generative learning theory (Fiorella & Mayer, 2015), we hypothesized students who received practice questions during the lesson would perform better on retention and transfer tests than students who did not receive practice questions (prediction 5). Predictions 4 and 5 also represent new contributions beyond the study by Parong and Mayer (2018).

2 | METHOD

2.1 | Participants and design

Sixty-one participants (40 female, 20 male and 1 other) aged 18–38 years (M = 20.51, SD = 3.27) were recruited through a recruitment website in the psychology department of a university in southern California. Participants were compensated 20 dollars for completion of the experiment. A power analysis was conducted using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) based on the effect sizes found in Parong and Mayer (2018). With an effect size of d = 1.12, power of 0.80, and alpha level of .05, 10 participants were needed in each group to detect an effect. In a 2 (IVR vs. PPT) × 2 (practice testing vs. no practice testing) between-subjects design, participants were randomly

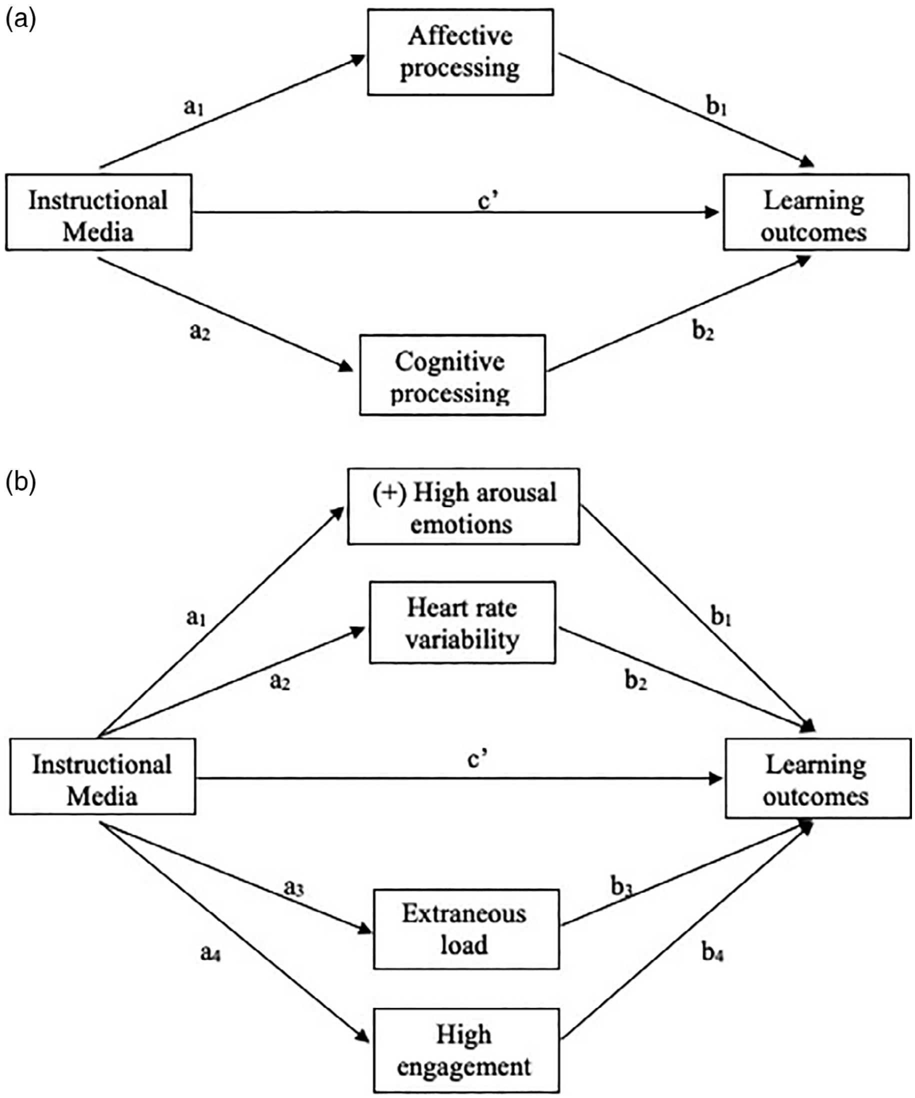

FIGURE 2 Proposed cognitive-affective distraction model of virtual reality learning environments. These models show a general (1a) and specific (1b) proposed cognitive-affective distraction models of immersive virtual reality learning environments, such that instructional media cause differences in cognitive and affective processing, which cause increases or decreases in learning outcomes

assigned to each of the four conditions: 15 participants each served in the IVR-practice testing (IVRp) group, IVR-no practice testing (IVRn) group and PPT-practice testing (PPTp) group, as well as 16 served in the PPT-no practice testing (PPTn) group.

2.2 | Materials

The computer-based materials included two versions of a biology lesson about the human bloodstream and the paper-based materials included surveys and post-tests. The biology lesson was displayed in either an immersive virtual reality simulation (IVR lesson) or a PowerPoint slideshow (PPT lesson). The IVR lesson, called The Body VR: Journey into a Cell, was a simulated lesson through the human bloodstream and into a cell (Figure 3; The Body VR LLC, 2016). In the IVR lesson, participants stood on a moving platform, which carried them through a narrated tour of the parts and functions of a blood vessel and a cell. Participants could occasionally interact with the elements they encountered; for example, a close up of a red blood cell appeared in front of the participant and the participant could touch and move it, as well as create more red blood cells using the controller. The objective of the lesson was to increase the learner's declarative knowledge of the roles and relationships of specifics components of the blood vessel (i.e., blood cells) and within a cell. The lesson lasted approximately 12 min and ran either continuously (for the IVRn group) or was displayed in six segments of approximately 2 min each (for the IVRp group). In the IVRp condition, the experimenter posed an oral stated question and waited for an orally stated answer from the participant after each of the six segments (e.g., “Which blood cell carries oxygen?”). The practice questions were not intended to give the participant any additional information beyond what was in the lesson. Participants were given unlimited time to answer, and no feedback was given after each answer.

The PPT lesson consisted of a set of 24 PowerPoint slides, as shown in Figure 4, which contained screenshots from the IVR lesson and corresponding printed text transcribed from the narration in the

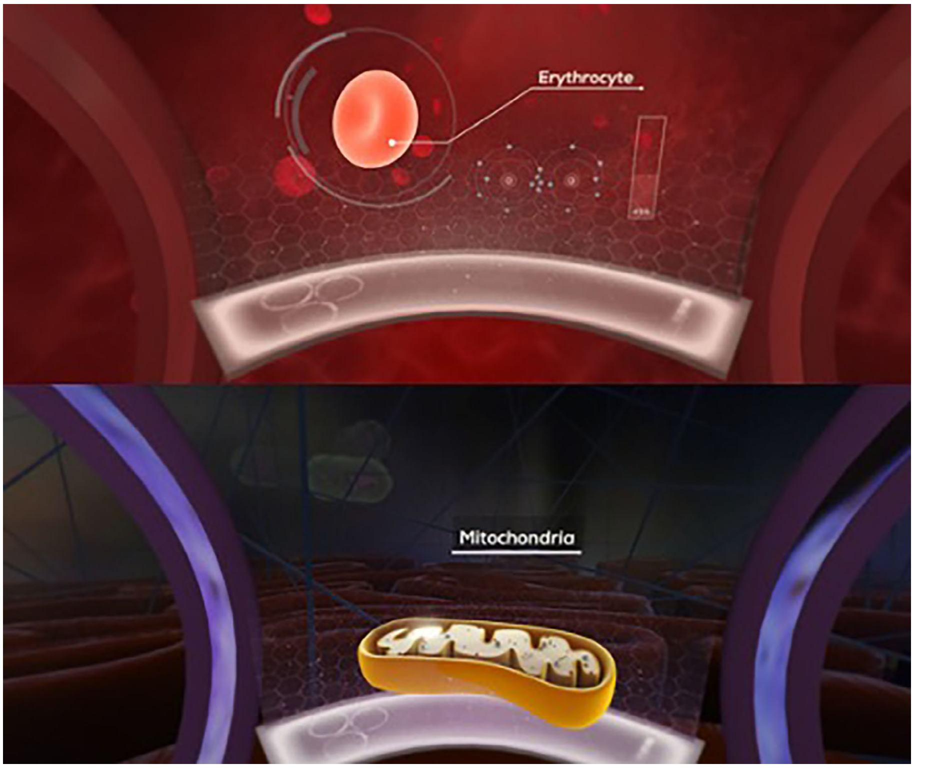

FIGURE 3 Screenshots from The Body VR: Journey Inside a Cell (The Body VR LLC, 2016) [Colour figure can be viewed at wileyonlinelibrary.com]

IVR lesson. The PPT slideshow was self-paced, and participants were given unlimited time to view each slide but could not return to previous slides. Participants took an average of 8 min to complete the PowerPoint slideshow. The slideshow ran continuously (for the PPTn group) or was displayed in six segments of two to seven slides each (for the PPTp group). In the PPTp version, there was an orally stated question after each of the six segments, identical to the IVRp group.

The paper-based materials included the pre-lesson questionnaire, post-lesson questionnaire and post-test. The pre-lesson questionnaire solicited demographic information, such as age, gender, year in school and major, and asked participants to indicate their background academic experience through a course survey (identifying science classes taken in high school or college), science-related activity checklist (e.g., "I have participated in science fairs"), and human body knowledge rating scale from 1 (very low) to 5 (very high). The background academic survey, activity checklist and rating scale were used to create a learner's prior knowledge score, which was used to verify that groups did not differ in prior knowledge as it predicts performance on learning outcomes. A pre-test on the content of the biology lesson was not used because it may have directed the participant's attention only to specific facts during the lesson and created a testing effect (Brown et al., 2014; Dunlosky et al., 2013).

The post-lesson questionnaire asked participants to rate a set of statements on a seven-point Likert-type scale from 1 (strongly disagree) to 7 (strongly agree). Sub-scales included presence, which had items modified from Witmer and Singer's (1998) presence questionnaire (7 items; e.g., "I felt like I was immersed (or included in) and interacting with the computer environment," Cronbach's α = .77), difficulty (2 items; e.g., "The lesson was difficult," α = .62), understanding (2 items; e.g., "I have a good understanding of the material," α = .78), effort, (2 items; e.g., "I put a lot of effort in the lesson", α = .91), interest (2 items; e.g., "I would like to see more of this type of lesson in classrooms," s α = .74), enjoyment (2 items; e.g., "I enjoyed the lesson," α = .93), and motivation (2 items; e.g., "I felt motivated to understand the lesson," α = .77). In addition, a set of questions measured the participant's perceived cognitive load, which included extraneous load (3 items; e.g., "I felt distracted during the lesson," α = .91), essential load (3 items; e.g., "I was trying to learn the main facts from the lesson," α = .64), and generative load (3 items; e.g., "I was trying to make connections between the material and things I already know," α = .66). Eight items asked about affective states: positive valence/high arousal (2 items; e.g., "I felt happy," α = .91), negative valence/high arousal (2 items; e.g., "I felt angry," α = .56), positive valence/low arousal (2 items; e.g., "I felt content," α = .59) and negative valence/low arousal (2 items; e.g., "I felt bored," α = .28).

The post-test consisted of 16 retention questions in the form of multiple-choice items (e.g., "In what process are ribosomes involved? (a) ATP production, (b) protein synthesis, (c) protein transportation, (d) B and C") and four transfer questions (e.g., "Describe what happens on a cellular level when you get a cut on your finger."), which were open-ended and had multiple correct answers. The retention questions were scored with 1 point for a correct answer and 0 points for an incorrect answer, yielding a total possible retention score of 16 (Cronbach's α = .77). Each of the four transfer questions was worth up to 3 or 4 points (i.e., 1 point for each possible correct answer), yielding a total possible transfer score of 15 (Cronbach's α = .65).

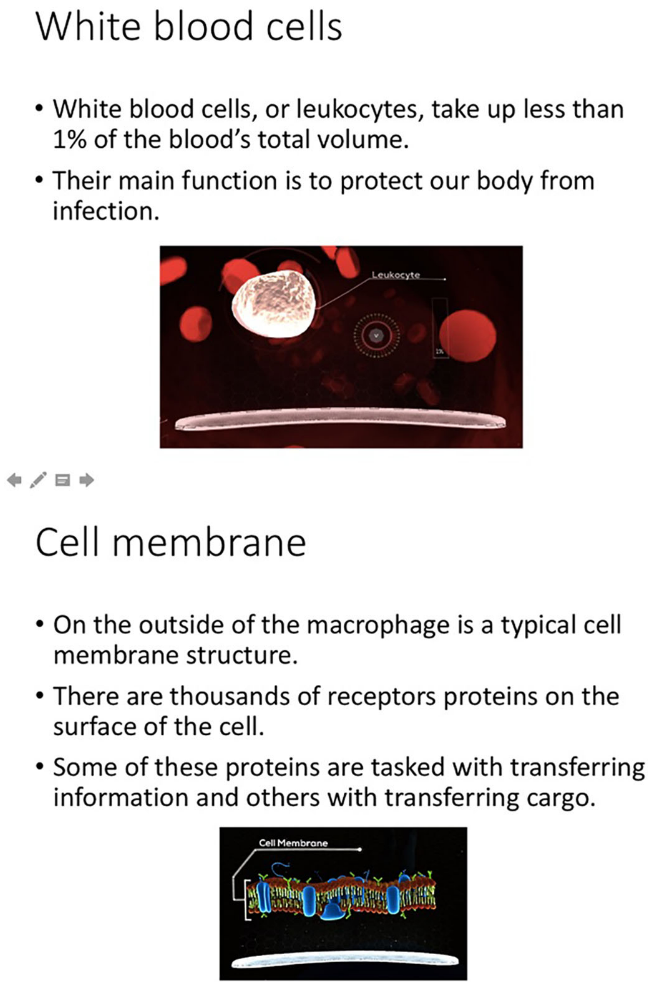

FIGURE 4 PowerPoint adapted from The Body VR: Journey Inside a Cell [Colour figure can be viewed at wileyonlinelibrary.com]

2.3 | Apparatus

The apparatuses included a virtual reality system, desktop computer system, brain-monitoring system and biometric-monitoring system. The virtual reality console was the HTC Vive and was used with Steam software to present the IVR lesson. The HTC Vive included a head-mounted display and a hand controller that the participant used to interact with the lesson. Two sensors placed in opposite corners of an 18 × 18-ft lab room were used to track the participant's movements within the lesson. The desktop computer system consisted of a Dell Alienware desktop computer with a 24-in. monitor.

The brain-monitoring system was Advanced Brain Monitoring's B-Alert X10 mobile EEG system, which included nine electrodes that corresponded to brain regions, Fz, F3, F4, Cz, C3, C4, POz, P3 and P4 (https://www.advancedbrainmonitoring.com/xseries/x10). It also included two leads that were attached to the collar bones to measure the electrocardiography (ECG) signal, which was used to calculate HR and HRV. To collect EDA, Biopac's wireless BioNomadix PPG/EDA module with MP160 system was used (https://www.biopac.com/product/bionomadix-ppg-and-eda-amplifier). A wireless transmitter was secured on the participant's wrist with a Velcro strap. Two disposable electrode pads with adhesive were placed on the palm near the thumb and pinky fingers, which were connected via wires to the transmitter.

The software used to acquire, process and analyse the EEG, ECG and EDA data was Biopac's AcqKnowledge version 5.0.3 with the Cognitive Brain State add-on (https://www.biopac.com/product/acqknowledge-software). Two EEG-based proprietary measures, cognitive workload and engagement, were calculated as measures of cognitive processing. The workload metric measured the participant's cognitive workload on a scale from 0 to 1, with measurements closer to 1 reflecting higher workload. It classified the participant's cognitive workload as low workload or boredom (<.40), manageable or ideal workload (.40–.70) and high workload or cognitive overload (>.70), which correlate with increased working memory load and difficulty level ratings in mental arithmetic and other complex problem-solving tasks (Poythress et al., 2006). These classifications have been used previously to measure workload during academic lessons (e.g., Makransky et al., 2017). The engagement metric classified the probability from 0 to 1 of the participant's current cognitive state into four classes, which included sleep onset, distraction, low engagement and high engagement. Three benchmark tasks, which included the short versions of the three-choice vigilance task (3CVT), visual psychomotor vigilance task (VPVT) and auditory psychomotor vigilance task (APVT) were used to calibrate the engagement metrics, which mapped on the participant's current brain state to one of the brain states listed above.

The EDA signal was used to calculate skin conductance level from the tonic EDA signal (SCL), skin conductance responses per second (SCR), which was the number of non-specific responses to stimuli per second averaged over a period of time (i.e., the duration of the lesson), and peak-to-peak (PTP) amplitude change from the phasic EDA signal, which indicated the change from the lowest and highest points in the phasic signal. These three measures were used to indicate physiological arousal with higher values reflecting higher physiological arousal. HR and HRV were calculated from the ECG signal. High-frequency HRV in the 0.15–0.40 Hz range was used as a measure of parasympathetic activity, with lower scores indicating more physiological arousal, particularly associated with stress, panic, anxiety and worry.

2.4 | Procedure

Participants were randomly assigned to one of four conditions: IVRn, IVRp, PPTn or PPTp. First, the participant gave his or her written consent to participate in the experiment. Next, the participant completed the pre-lesson questionnaire at his or her own pace. Then, the electrodes for the EEG, EDA and ECG systems were set up on the participant by the researcher. The participant was then given instructions on his or her assigned lesson. Participants in the IVRn and IVRp conditions were moved to the middle of the room while a researcher set up the head-mounted display on the participant and gave the participant instructions on using the controller for the lesson. Those in the IVRp condition were given additional instructions about the practice questions; they were told that the researcher would occasionally stop the lesson and ask a question out loud, and the participants would be given unlimited time to answer out loud. Participants in the PPTn and PPTp conditions were seated in front of a computer and given instructions on how to advance the slides in the lesson. They were told that they would view the lesson once at their own pace, and they could not return to the previous slide. Those in the PPTp condition were given additional instructions that there would be occasional screens with the word "STOP," during which the researcher would ask them a question out loud and wait for a response from the participant.

Before starting the lesson, a 2-minute baseline recording of the EEG, ECG and EDA data was collected. As EDA data are particularly sensitive to body movements, the number of large movements involving the limbs, head or trunk (e.g., turning body, taking a step within the environment) other than the minimal movement to interact with the lessons (i.e., extension of the arm forward used for reaching for the keyboard/mouse or pointing the IVR controller forward to make a selection) were counted to attempt to control for movements' effects on EDA data. After completing the lesson, the participants completed the post-lesson questionnaire and post-test. IRB approval was obtained and guidelines for research with human subjects were followed.

3 | RESULTS

3.1 | Preliminary analyses

We first determined whether there were pre-existing differences among groups despite random assignment. A chi-square analysis revealed that the proportion of men and women did not differ significantly among the four groups, $\chi^2(6, N = 61) = 3.45$, $p = .751$. Analyses of variance (ANOVAs) showed that the groups did not differ significantly in mean age, $F(3, 57) = 1.63$, $p = .194$, or prior knowledge score, $F(3, 57) = 0.64$, $p = .592$. However, the groups differed significantly in the year in school, $F(3, 57) = 2.95$, $p = .040$. Post-hoc analyses showed that the PPTp group was significantly higher in the year in school than the IVRn group ($p = .028$) and IVRp group ($p = .024$). Therefore, year in school was added as a covariate in learning outcome analyses to control for pre-existing differences among groups.

Main effects of instructional media (i.e., IVR or PPT) and instructional methods (i.e., practice testing or no practice testing) based on 2 ×2 ANOVAs and analyses of covariance (ANCOVAs) are reported in the following sections according to the order of predictions. Estimated marginal means (EMMs) are reported for ANCOVAs. Although no predictions were made about interactions between instructional media and practice testing, they were examined, and results are reported along with the predictions regarding instructional media and methods. Due to a high number of prior multiple comparisons, significances of p values were evaluated using the Benjamini–Hochberg procedure, controlling for a false discovery rate of 5% (Benjamini & Hochberg, 1995).

3.2 | Instructional media

3.2.1 | Did instructional media affect learning outcomes?

According to the idea that IVR creates distraction, the first prediction is that students who learn with a PowerPoint slideshow will score higher on the retention and transfer tests than students who learn in IVR. Table 1 shows the mean retention score (and standard deviation) and mean transfer test score (and standard deviation) for each group. Separate 2 (IVR vs. PPT) × 2 (practice testing vs. no practice testing) ANCOVAs were run on retention test scores and on transfer test scores, with the year in school as a covariate. On retention scores, the PPT groups (EMM = 10.78, $SE = .63$) scored higher than the IVR groups (EMM = 8.99, $SE = .64$), $F(1, 56) = 3.75$, $p = .058$, $d = -0.52$, but the difference did not reach statistical significance. On transfer scores, the PPT groups (EMM = 3.76, $SE = .30$) scored significantly higher than the IVR groups (EMM = 2.57, $SE = .31$), $F(1, 56) = 7.11$, $p = .010$, $d = -0.71$. There were no significant interactions between instructional media and practice testing on retention scores, $F(1, 56) = 0.34$, $p = .565$, or transfer scores, $F(1, 56) = 0.04$, $p = .843$. Overall, prediction 1 was partially supported, particularly for transfer test scores.

3.2.2 | Did instructional media affect affective processes?

According to the distraction hypothesis, the PPT groups are expected to give higher ratings of positive, high emotional arousal (prediction 2a) and score higher on biometric measures of arousal (prediction 2b). Table 2 shows mean emotional arousal ratings (and standard deviations) for each group. ANOVAs were run on ratings of each of the four self-reported affective states with instructional media (IVR vs. PPT) and practice testing (practice testing vs. no practice testing) as between-subjects factors. There was a significant instructional media effect for the positive valence/high arousal state and the negative valence/low arousal state, in which the IVR conditions ($M = 5.35$, $SD = 1.36$) reported a more positive and highly arousing affective state than those in the PPT conditions ($M = 3.37$, $SD = 1.35$), $F(1, 57) = 33.94$, $p < .001$, $d = 1.54$, and the PPT conditions ($M = 2.26$, $SD = 1.06$) reported a more negative and low arousing state than the IVR conditions ($M = 1.52$, $SD = 0.75$), $F(1, 57) = 9.61$, $p = .003$, $d = -0.82$. There was no significant difference between the IVR ($M = 1.18$, $SD = 0.43$) and PPT conditions (M = 1.36, $SD = 0.69$) on the negative/high arousal measures, $F(1, 57) = 1.59$, $p = .213$, $d = -0.33$, and there was no difference between IVR ($M = 4.93$, $SD = 1.06$) and PPT conditions ($M = 4.94$, $SD = 1.20$) on the positive/low arousal measures, $F(1, 57) < 0.01$, $p = .967$, $d < 0.001$.

There was a significant interaction between instructional media and practice testing for the positive valence/low arousal state, such that practice testing in the IVR conditions increased positive and low arousing emotions during the lesson, while in the PPT conditions,

TABLE 1 Means (and *SD*) on learning outcome scores by group

	Retention score		Transfer score	
	IVR M (SD)	PPT M (SD)	IVR M (SD)	PPT M (SD)
No practice	8.53 (3.25)	10.75 (3.92)	2.13 (1.36)	3.53 (1.85)
Practice	9.53 (2.53)	10.73 (3.43)	2.87 (1.25)	4.13 (1.89)

Abbreviations: IVR, immersive virtual reality; PPT, PowerPoint slideshow.

TABLE 2 Means (and *SD*) of self-reported emotional states during the lesson for all groups

	Positive, high arousal		Positive, low arousal		Negative, high arousal		Negative, low arousal	
	IVR	PPT	IVR	PPT	IVR	PPT	IVR	PPT
No practice	5.70 (1.13)	3.66 (1.14)	4.67 (1.25)	5.34 (0.77)	1.17 (0.41)	1.13 (0.34)	1.53 (0.83)	2.19 (0.89)
Practice	5.00 (1.50)	3.07 (1.52)	5.20 (0.77)	4.50 (1.43)	1.20 (0.46)	1.60 (0.87)	1.50 (0.68)	2.33 (1.25)

Note: Each averaged score was on a scale from 1 to 7, with higher scores indicating increased emotion.
Abbreviations: IVR, immersive virtual reality; PPT, PowerPoint slideshow.

TABLE 3 Means (and *SD*) of electrodermal activity measures during the lesson for all groups

	SCL		SCR/sec		Peak-to-peak	
	IVR	PPT	IVR	PPT	IVR	PPT
No practice	10.77 (5.62)	9.90 (6.34)	.09 (.03)	.06 (.03)	3.93 (2.70)	2.95 (2.57)
Practice	12.95 (7.92)	9.56 (4.38)	.08 (.03)	.09 (.03)	7.05 (4.12)	4.48 (2.92)

Note: EDA was measured in microsiemens. SCL is the average skin conductance level from the tonic signal. SCR/sec is the non-specific skin conductance responses per second. Peak-to-peak is the difference between highest and lowest points of phasic EDA signal.
Abbreviations: IVR, immersive virtual reality; PPT, PowerPoint slideshow.

practice testing decreased these emotions, $F(1, 57) = 6.09$, $p = .017$. Interactions between instructional media and practice testing on positive/high arousal emotions, $F(1, 57) = 0.03$, $p = .872$, negative/high arousal emotions, $F(1, 57) = 2.41$, $p = .126$, and negative/low arousal emotions, $F(1, 57) = 0.14$, $p = .710$, were not significant. Overall, these findings support prediction 2a and suggest that learning in virtual reality is emotionally arousing, particularly with positive emotions.

According to the distraction hypothesis, the PPT groups are expected to score higher on EEG-based measures of arousal (prediction 2b). Table 3 shows the means and standard deviations on three measures of EDA, and Table 4 shows the mean and standard deviations on two measures of heart activity from electrocardiography (ECG). ANCOVAs were conducted on these measures with media (IVR vs. PPT) and practice testing (practice testing vs. no practice testing) as between-subjects factors. The covariates used were the number of large movements the participants made during the lesson and the respective baseline measure for each variable.

For the EDA measures, the IVR groups (EMM = 10.66, $SE = .42$) and PPT groups (EMM = 10.95, $SE = .42$) did not differ significantly on SCL, $F(1, 55) = 0.23$, $p = .637$, $d = -0.13$. There was also no significant difference between IVR groups (EMM = .08, $SE = .004$) and PPT groups (EMM = .08, $SE = .004$) on SCR, $F(1, 55) = 0.17$, $p = .678$, $d = 0.11$; and the IVR groups (EMM = 5.23, $SE = .56$) did not have significantly different PTP amplitude change compared with the PPT groups (EMM = 3.98, $SE = .55$), $F(1, 55) = 2.37$, $p = .130$, $d = 0.41$. There were no significant interactions (with the Benjamini–Hochberg correction for false discovery rate) between instructional media and practice testing on SCL, $F(1, 55) = 4.20$, $p = .045$, SCR/sec, $F(1, 55) = 5.66$, $p = .021$, or PTP amplitude, $F(1, 55) = 0.34$, $p = .561$.

The IVR groups (EMM = 6.23, $SE = 7.70$) did not differ significantly on HRV than the PPT groups (EMM = 13.37, $SE = .38$), $F(1, 54) = 0.38$, $p = .543$, $d = -0.17$. The IVR groups (EMM = 84.48, $SE = .90$) and PPT groups (EMM = 82.60, $SE = .89$) also did not differ significantly on HR, $F(1, 55) = 1.99$, $p = .165$, $d = 0.38$. There were no interactions involving instructional media and practice testing on HRV, $F(1, 54) = 0.30$, $p = .586$, or heart rate, $F(1, 55) = 3.29$, $p = .075$. These findings do not support prediction 2b; learning in IVR did not create more affective processing as measured by physiological arousal than learning with conventional media.

3.2.3 | Did instructional media affect cognitive processes?

Moreover, according to the distraction hypothesis, the PPT groups are expected to score lower than the IVR groups on ratings of extraneous cognitive load (prediction 2a). Table 5 shows the mean cognitive load

TABLE 4 Means (and *SD*) of electrocardiography measures during the lesson for all groups

	Heart rate		Heart rate variability	
	IVR	PPT	IVR	PPT
No practice	87.60 (8.17)	79.45 (7.62)	22.47 (80.28)	1.02 (1.60)
Practice	88.46 (9.73)	78.71 (11.01)	10.05 (23.69)	5.66 (10.09)

Note: Heart rate was measured in beats per minute (BPM). Heart rate variability (HRV) was measured using the high-frequency band (0.15–0.40 Hz, ms^2).
Abbreviations: IVR, immersive virtual reality; PPT, PowerPoint slideshow.

TABLE 5 Means (and Standard Deviations) of self-reported cognitive load scores during the lesson by group

	Extraneous load		Essential load		Generative load	
	IVR	PPT	IVR	PPT	IVR	PPT
No practice	3.33 (1.85)	2.35 (1.06)	4.82 (1.27)	5.04 (1.02)	5.22 (1.45)	5.40 (0.82)
Practice	3.71 (1.46)	2.44 (1.36)	5.71 (0.87)	5.27 (0.98)	5.33 (0.64)	4.78 (1.38)

Note: Each averaged score was on a scale from 1–7, with higher scores indicating increased cognitive load.
Abbreviations: IVR, immersive virtual reality; PPT, PowerPoint slideshow.

TABLE 6 Means (and *SD*) of EEG-based cognitive metrics during the lesson by group

	High engagement		Low engagement		Distraction		Sleep onset		Workload	
	IVR	PPT	IVR	PPT	IVR	PPT	IVR	PPT	IVR	PPT
No practice	.34 (.17)	.43 (.15)	.46 (.20)	.38 (.12)	.08 (.17)	.07 (.12)	.10 (.11)	.08 (.12)	.43 (.19)	.50 (.15)
Practice	.42 (.19)	.40 (.12)	.39 (.25)	.44 (.21)	.04 (.07)	.05 (.06)	.11 (.14)	.09 (.13)	.53 (.16)	.51 (.13)

Note: The high engagement, low engagement, distraction and sleep onset cognitive engagement metrics were indicated the probability of the learner being in that cognitive state. The workload metric was on a scale from 0 to 1, with .0–.39 indicating boredom, .40–.69 indicating ideal cognitive workload, and .70–1.00 indicating cognitive overload.
Abbreviations: IVR, immersive virtual reality; PPT, PowerPoint slideshow.

ratings (and standard deviations) for each group. As predicted, the IVR groups (M = 3.52, SD = 1.65) reported significantly higher extraneous cognitive load than the PPT groups (M = 2.40, SD = 1.20), F(1, 57) = 9.08, p = .004, d = 0.80. However, the IVR groups (M = 5.27, SD = 1.16) and the PPT groups (M = 5.15, SD = 0.98) did not differ in essential cognitive load ratings, F(1, 57) = 0.18, p = .675, d = 0.11, and the difference between the generative cognitive load ratings by the IVR groups (M = 5.28, SD = 1.10) and the PPT groups (M = 5.10, SD = 1.15) was also not significant, F(1, 57) = 0.44, p = .509, d = 0.18. There were no significant interactions between instructional media and practice testing on perceived extraneous cognitive load, F(1, 57) = 0.15, p = .701, essential load, F(1, 57) = 1.55, p = .218 or generative load, F(1, 57) = 1.61, p = .210. These findings partially support prediction 2a, particularly with respect to reported extraneous load, which is most diagnostic of distraction.

The PPT groups are expected to score higher on EEG-based measures of engagement and ideal cognitive load (prediction 2b). Table 6 shows mean EEG measures of engagement and cognitive load (and standard deviations) for each group. Two (IVR vs. PPT) × two (practice testing vs. no practice testing) ANCOVAs were conducted on workload scores (i.e., level of workload) and engagement scores (i.e., probability of being in a certain state of cognitive engagement relative to other states), with respective baseline measures as covariates. As predicted, students in the PPT conditions (EMM = .47, SE = .03) exhibited a higher probability of being in a high engagement cognitive state than students in IVR conditions (EMM = .33, SE = .03), F(1, 56) = 13.66, p < . 001, d = –0.99. There were no significant differences between IVR (EMM = .41, SE = .02) and PPT groups (EMM = .42, SE = .02) on probabilities of being in a low engagement state, F(1, 56) = 0.04, p = .851, d = –0.06. IVR (EMM = .09, SE = .02) and PPT groups (EMM = .04, SE = .02) also did not differ in probabilities of being in a distraction state, F(1, 56) = 2.69, p = .107, d = 0.44. Finally, IVR (EMM = .10, SE = .02) and PPT groups (EMM = .09, SE = .02) did not differ in their probabilities of sleep onset state, F(1, 56) = 0.30, p = .585, d = 0.14. There were no significant interactions between instructional media and practice testing on high cognitive engagement, F(1, 57) = 0.18, p = .674, low cognitive engagement, F(1, 57) = 0.38, p = .542, distraction, F(1, 57) = 0.13, p = .720 or sleep onset, F(1, 57) = 0.03, p = .857.

Comparing the types of cognitive workload the learners exhibited during the lesson, chi-square analyses revealed that the PPT groups had a higher proportion of participants with ideal cognitive workload (compared to boredom or cognitive overload) during the lesson than the IVR groups, χ^2(1, N = 61) = 6.34, p = .012, Cramer's V = 0.32. In addition, the IVR groups had a higher proportion of bored learners than PPT, χ^2(1, N = 61) = 5.79, p = .012, Cramer's V = .31. However, the groups did not differ in their proportions of learners with cognitive overload, $\chi^2(1, N = 61) = 0.26$, $p = .614$, Cramer's $V = 0.07$. Table 7 shows the frequencies of each type of cognitive workload learners in each condition experienced during the lesson. Overall, these findings partially support prediction 3b; there is evidence that learning in IVR creates more distraction during learning than learning with a desktop.

3.2.4 | Did cognitive or affective processes mediate the relationship between instructional media and learning outcomes?

Prediction 4 is that cognitive and affective measures will mediate the relationship between the level of immersion during instruction (i.e., PPT versus IVR) and test performance (i.e., on retention and transfer tests). A parallel mediation was conducted, as shown in Figure 2a. Extraneous cognitive load and high engagement were chosen as measures of cognitive processing to determine whether the amount of distraction the learner perceived or the amount of engagement during the lesson mediated this effect. Self-reported positive/high arousal emotions and HRV were selected as measures of affective processing to determine whether distraction due to positive emotional arousal or physiological arousal mediated this relationship. Although EDA and HRV both measure physiological arousal, HRV was chosen as it was less sensitive to the movements that may have been confounded between the two types of media. Figure 2b shows this updated model. The mediation analysis was conducted with the PROCESS macro in SPSS (Hayes, 2018) based on 5,000 bootstrapped samples. Instructional media was coded as a categorical variable with the IVR lesson as 0 and the desktop lesson as 1.

In summary, Figure 5 shows the two mediation models run on retention and transfer scores. The total effect of instructional media on retention scores was significant ($b = -1.93$, $SE = .83$, $t = -2.34$, $p = .023$), and the direct effect of instructional media on retention scores after including the mediators was also significant ($b = -2.39$, $SE = 1.07$, $t = -2.32$, $p = .030$). The indirect effect of instructional media through self-reported extraneous load was significant ($b = -0.78$, $SE = .02$, 95% CI [−1.73, −.18]). Those in the IVR conditions reported higher extraneous load, which caused lower retention scores than the PPT conditions. The indirect effect of instructional media through self-reported positive/high arousal emotions was also significant ($b = 1.22$, $SE = .67$, 95% CI [0.04, 2.71]). Those in the IVR conditions reported more positive/high arousal emotions, which caused lower retention scores than the PPT conditions, overall. The indirect effects through high engagement ($b = -0.14$, $SE = .22$, 95% CI [−0.69, −0.17]) and HRV ($b = 0.16$, $SE = .15$, 95% CI [−0.03, 0.53]) were not significant.

For transfer scores, the total effect of instructional media was significant ($b = -1.45$, $SE = .40$, $t = -3.61$, $p < .001$), and the direct effect of instructional media on transfer scores after including the mediators was reduced, but still significant ($b = -1.53$, $SE = .57$, $p = .010$). The indirect paths of instructional media on transfer scores through extraneous cognitive load ($b = -0.16$, $SE = .16$, 95% CI [−0.52, 0.13]), high engagement (b = −0.07, $SE = .10$, 95% CI [−0.28, 0.14]), positive, high arousal emotions ($b = 0.32$, $SE = .36$, 95% CI (−0.36, 1.07]) and HRV ($b = -0.01$, $SE = .06$, 95% CI [−0.11, 0.16]) did not significantly mediate the relationship. These results partially support the hypothesis that cognitive and affective processes, particularly self-report measures, mediate the relationship between instructional media and learning outcomes, but only for retention tests.

TABLE 7 Frequencies of type of cognitive workload during the lesson for all groups

		Cognitive workload type		
		Bored	**Ideal**	**Overload**
Group				
IVR	No practice	9	6	0
	Practice	3	9	3
PPT	No practice	2	13	1
	Practice	2	12	1

Abbreviations: IVR, immersive virtual reality; PPT, PowerPoint slideshow.

3.3 | Practice testing

3.3.1 | Did practice testing affect learning?

Contrary to the predictions of the generative learning strategy hypothesis (prediction 5), as summarized in Table 1, scores on the retention test did not differ significantly for students who engaged in practice testing (EMM = 10.14, $SE = .61$) versus students who did not (EMM = 9.64, $SE = .60$), $F(1, 56) = 0.34$, $p = .565$, $d = 0.16$; and scores on the transfer test did not differ significantly for students who engaged in practice testing (EMM = 3.50, $SE = .30$) versus students who did not (EMM = 2.84, $SE = .29$), $F(1, 56) = 2.49$, $p = .120$, $d = 0.42$.

3.4 | Did instructional media or practice testing affect presence or other self-report measures?

Multiple 2 (IVR vs. PPT) × 2 (practice testing vs. no practice testing) ANOVAs were run on self-reported ratings of presence, difficulty, understanding, effort, interest, enjoyment and motivation. Results revealed that the IVR groups ($M = 5.12$, $SD = 0.80$) reported significantly higher presence during the lesson than the PPT groups ($M = 3.57$, $SD = 1.01$), $F(1, 57) = 43.10$, $p < .001$, $d = 1.74$. The IVR groups ($M = 5.83$, $SD = 1.54$) reported significantly higher enjoyment during the lesson than the PPT groups ($M = 3.98$, $SD = 1.64$), $F(1, 57) = 20.59$, $p < .001$, $d = 1.20$. No other comparisons between media significantly differed in perceived difficulty, understanding, effort, interest and motivation. No main effects of practice testing and no interactions between media and practice testing on any of these self-report measures were significant. As expected, students felt more presence, enjoyed the experience more and perceived the lesson as

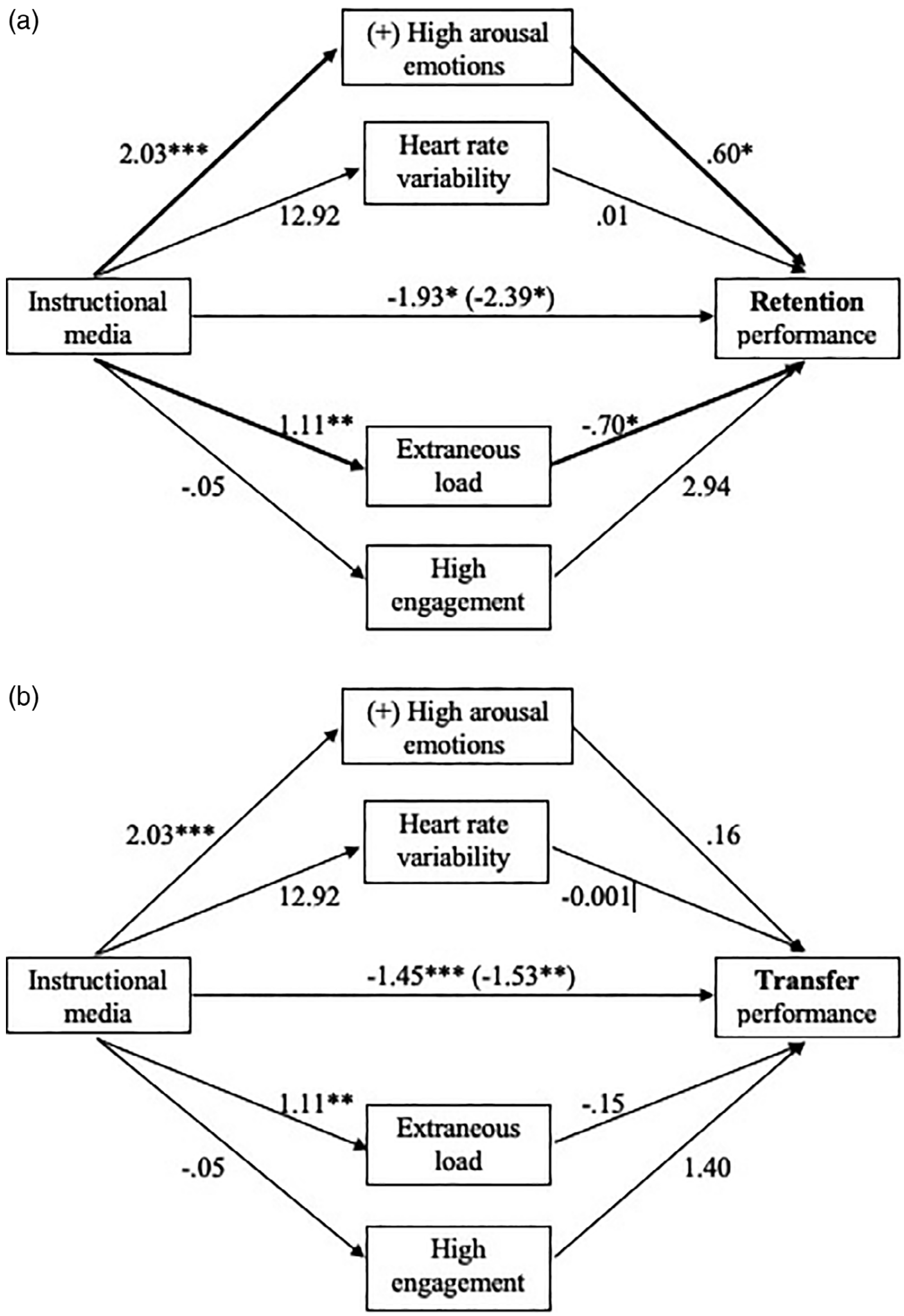

FIGURE 5 Mediation models predicting learning outcomes. These models explain the relationship between instructional media and (a) retention performance and (b) transfer performance for a science lesson. Instructional media was coded as a categorical variable with the PPT condition as a 0 and the IVR condition as a 1. Coefficients represent unstandardized regression coefficients. Coefficients in parentheses (c′) indicate the direct effect of instructional media on learning after controlling for the mediators. Bold lines indicate significant indirect effects (i.e., mediation). *$p < .05$, **$p < .01$, ***$p < .001$

easier in IVR than with conventional media; however, liking the experience did not translate into better learning processes or outcomes.

4 | DISCUSSION

4.1 | Empirical implications

The primary purpose of this study was to examine the differences in academic learning between an immersive lesson presented in IVR and a more traditional lesson delivered via a desktop computer. Table 8 displays a summary of these findings. Learners performed better on transfer tests after viewing a biology lesson in a PowerPoint slideshow as compared to an equivalent lesson in IVR. IVR was more cognitively distracting based on self-report measures and EEG-based measures than the PowerPoint; students who viewed the PowerPoint slideshow reported less extraneous cognitive load and showed higher probabilities of being in a high cognitive engagement state, and there are more learners with ideal cognitive workload than the IVR group. IVR also led to higher reports of high arousal, positive emotions. Mediation analyses showed that measures of distraction, particularly self-reported extraneous load and self-reported positive, high arousal emotions, partially explain the relationship between instructional media and learning outcomes in the form of retention tests. Overall,

TABLE 8 Summary of the effects of media on each outcome measure

	Media	
Outcome measure	IVR	PPT
Affective processing		
Self-reported emotions		
Positive/high arousal	+	–
Negative/low arousal	–	+
Positive/low arousal	No sig. Difference	
Negative/high arousal	No sig. Difference	
Physiological measures		
All EDA measures	No sig. Difference	
All heart rate measures	No sig. Difference	
Cognitive processing		
Self-reported cognitive load		
Extraneous cognitive load	+	–
Essential cognitive load	No sig. Difference	
Generative cognitive load	No sig. Difference	
EEG measures		
High engagement	–	+
Ideal workload	–	+
Learning outcomes		
Retention	No sig. Difference	
Transfer	–	+

the results support the distraction hypothesis; in that IVR led to more cognitive distraction, which was associated with poorer performance on learning outcomes. These results are in line with previous empirical studies, that found that academic lessons displayed in IVR led to worse learning outcomes than other media (Makransky et al., 2017; Moreno & Mayer, 2004; Parong & Mayer, 2018), as well as a related meta-analysis that found that IVR produced slightly worse learning outcomes from social skill interventions than non-immersive displays (Howard & Gutworth, 2020). This adds to the scant body of work examining learning outcomes from the use of IVR.

4.2 | Theoretical implications

The results are consistent with the CTML. The perceptual richness of IVR and the high arousal emotions associated with it can serve as seductive details (interesting, but irrelevant, material that is added to a lesson) that distract the learner from engaging in deep processing of the target content (Plass & Kalyuga, 2019). Specifically, the coherence principle states that people learn better when these extraneous details are removed from the lesson. The IVR lesson had extraneous sounds, animations and interactions that were not pertinent to learning the information from the lesson. These features, as well as the use of an unfamiliar technology, could have also induced extraneous cognitive or affective processing. Based on CTML, these irrelevant details in the lessons could have led to distracting learning by interrupting the selection (e.g., moving attention away from the content to irrelevant stimuli in the lesson or to the learner's emotional processing), organization (e.g., incorporating non-essential details or feelings of arousal into the learner's mental model of the lesson) and integration (e.g., relating the salient emotions to prior knowledge, rather than the content in the lesson) processes during learning. Because the IVR lesson violated the coherence principle by including extraneous stimuli or inducing extraneous emotion, some of the learner's cognitive resources would have been used to process these stimuli, thereby reducing cognitive resources for essential and generative workload. This is in line with previous research, examining the coherence principle in non-IVR environments (Mayer et al., 1996; Mayer, Heiser, & Lonn, 2001; Moreno & Mayer, 2000).

4.3 | Practical implications

A practical question is whether IVR should be used in the classroom. Based on the results of this study, using IVR to replace video lessons or other types of instructional media may not yet be useful, particularly for lessons similar to the one used in this study. However, some evidence has shown that IVR is useful for learning in some cases (e.g., Kozhevnikov et al., 2013, Webster, 2016). In addition, IVR technologies are currently used across several disciplines for training and education, including engineering, computer science, astronomy and nursing, and are often touted as at least promising in many fields (see Radianti et al., 2020, for review). Future research should determine moderating factors of the effectiveness of IVR lessons, such as the domain of the lesson or type of knowledge tested. For example, benefits of IVR could be domain specific, or more specifically, depend on whether the lesson involves the use of spatial abilities, such as a navigation or mental rotation. During these lessons, the use of a 3D learning environment, rather than a 2D environment that is meant to represent a 3D space, such as a desktop computer, may be useful. Immersing the learner in a 3D learning environment may reduce the cognitive resources needed to mentally transform a 2D image into 3D object or space, thereby facilitating learning. In addition, this study only examined the effects of IVR on declarative knowledge. IVR may have different effects on other types of knowledge, such as analytical, creative or evaluative thinking, which future research can address by adding multiple learning outcome measures.

In addition, because of the positive ratings by students and teachers of IVR lessons, IVR may be a useful venue for pre-training before conventional lessons rather than as a replacement for them. For example, they could be used to familiarize students with key words, concepts or images before a lecture. This would help the learner manage essential load during the lesson as he or she would spend less time processing an already-familiar term or picture. The pre-training principle has had robust effects on learning outcomes among other types of lessons (Mayer & Pilegard, 2014).

4.4 | Limitations and future directions

The practice testing effect was not replicated in this study, which could have been due to a number of reasons. First, practice testing effects may not be as apparent when measured with an immediate test, rather than a delayed test (Roediger & Karpicke, 2006). Second, the oral format in which it was delivered was different than the test format, which included written, multiple-choice and short answer questions. Practice testing effects may be more effective when the practice format is similar to the test format (e.g., written practice questions and a written retention test; Dunlosky et al., 2013). However, some research shows that practice tests can benefit learning even when they do not match the final test (e.g., using a multiple-choice practice test and a cued recall final test; Fazio, Agarwal, Marsh, & Roediger, 2010). One-on-one oral testing may have been stressful for students, thereby reducing its effectiveness. Third, the practice test included only one test question per section; practice testing may be more effective with more questions. Future research should examine how to implement practice testing in IVR in ways that are helpful for learning. Finally, the number of participants in the study was low. Although a power analysis determined the sample size to be large enough to detect differences in learning outcomes, there may not have been enough power to detect differences in practice testing.

Another limitation may be in the media comparison research itself. Media comparison studies are often critiqued because it is difficult to separate the effects of instructional media from the effects of instructional method (Clark & Feldon, 2014). For example, in this study, using IVR as the medium was confounded with the instructional method in which the lesson was displayed; IVR used continuous animation and spoken narration, whereas the PowerPoint slideshow contained static pictures and written text. In order to maximize experimental control, the same graphics and script were used to deliver equivalent academic content in both media. However, more work is needed to experimentally control and isolate the features of IVR that caused the observed effects.

Finally, it is possible that the results of the study may be explained by a novelty effect, in which the use of a new technology caused poorer learning outcomes and/or distraction. Many participants reported that it was their first time using an IVR console. The unfamiliarity of using the IVR interface, including wearing a head-mounted display and moving the controllers in a specific way, could have induced extraneous processing until the user had learned, and was comfortable with, the interactions with the interface. Cognitive resources would have been directed to counteract the feeling of wearing a headset or monitor the position of the controllers, leaving fewer cognitive resources for essential or generative processing of the incoming information. It would be important to determine whether this novelty effect could be diminished by comparing learning outcomes between habitual VR users and novice users, or using a training intervention to familiarize naïve learners with the mechanisms required to interact with the IVR interface. It is possible that as IVR proliferates classrooms and households in the coming years, this novelty effect would be diminished, and learning from IVR would become as familiar as learning with other multimedia technologies.

ACKNOWLEDGEMENTS

This project was supported by Grant N000141612046 from the Office of Naval Research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Jocelyn Parong https://orcid.org/0000-0001-7076-2535
Richard E. Mayer https://orcid.org/0000-0003-4055-6938

REFERENCES

Andreassi, J. L. (2007). Electrodermal activity (EDA) and behavior. In *Psychophysiology: Human behavior and physiological response* (5th ed., pp. 314–340). New York, NY: Psychology Press. https://doi.org/10.4324/9780203880340

Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology*, *10* (3), 229–240. https://doi.org/10.1037/1089-2680.10.3.229

Bailenson, J. (2018). *Experience on demand*. New York, NY: Norton.

Bailenson, J., & Blascovich, J. (2011). *Infinite reality: Avatars, eternal life, new worlds and the dawn of the virtual revolution*, New York, NY: Harper Collins.

Basar, E. (1999). Brain function and oscillations: Volume II: Integrative brain function. In *Neurophysiology and cognitive processes*, Berlin, Germany: Springer.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *Series B*, *5*(1), 289–300.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., … Olmstead, R. (2004). Real-time analysis of EEG indices of alertness, cognition, and memory with a wireless EEG headset. *International Journal of Human-Computer Interaction*, *17*(2), 151–170. https://doi.org/10.1207/s15327590ijhc1702_3

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., … Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, *78*(5) Suppl, B231–B244.

Bradley, M. M., & Lang, P. J. (2007). Emotion and motivation. In J. Cacioppo, L. G. Tassinary, & G. G. Bernston (Eds.), *The handbook of psychophysiology* (3rd ed., pp. 581–607). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

Brown, P. C., Rodiger, H. L. I. I. I., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*, Cambridge: MA: Harvard University Press.

Canli, T., Desmond, J. E., Zhao, Z., & Gabrieli, J. D. E. (2002). Sex differences in the neural basis of emotional memories. *Proceedings of the National Academy of Sciences*, *99*(16), 10789–10794. https://doi.org/10.1073/pnas.162356599

Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, *132*, 310–324. https://doi.org/10.1037/0096-3445.132.2.310

Clark, R. E., & Feldon, D. F. (2014). Ten common but questionable principles of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 151–173). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.009

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

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

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/bf03193146

Fazio, L. K., Agarwal, P. K., Marsh, E. J., & Roediger, H. L. (2010). Memorial consequences of multiple-choice testing on immediate and delayed tests. *Memory & Cognition*, *38*(4), 407–418. https://doi.org/10.3758/MC.38.4.407

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*, New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781107707085

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: a regression based approach*, (2nd ed).New York, NY: Guilford Press.

Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology*, *41*(1), 33–55. https://doi.org/10.1111/j.1467-8535.2008.00900.x

Howard, M. C., & Gutworth, M. B. (2020). A meta-analysis of virtual reality training programs for social skill development. *Computers & Education*, *144*(2020, 103707. https://doi.org/10.1016/j.compedu.2019.103707

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation*, *15*(3), 235–245. https://doi.org/10.30773/pi.2017.08.17

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, *42*(3), 419–427. https://doi.org/10.1007/bf02344719

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

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

LaValle, S. M. (2019). *Virtual reality*, Cambridge, England: Cambridge University Press.

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, *35*, 691–707. https://doi.org/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), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2017). 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

Markowitz, D. M., Laha, R., Perone, B. P., Pea, R. D., & Bailenson, J. N. (2018). Immersive virtual reality field trips facilitate learning about climate change. *Frontiers in Psychology*, *9*, 1–20. https://doi.org/10.3389/fpsyg.2018.02364

Mather, M., Mitchell, K. J., Raye, C. I., Novak, D. L., Greene, E. J., & Johnson, M. K. (2006). Emotional arousal can impair feature binding in working memory. *Journal of Cognitive Neuroscience*, *18*(4), 614–625. https://doi.org/10.1162/jocn.2006.18.4.614

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.009

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, *88*(1), 64–73. https://doi.org/10.1037/0022-0663.88.1.64

Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction*, *33*, 12–18.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.009

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187–198. https://doi.org/10.1037/0022-0663.93.1.187

Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 316–344). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.009

Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). Five popular study strategies: Their pitfalls and optimal implementations. *Perspectives on Psychological Science*, *13*, 390–407. https://doi.org/10.1177/1745691617710510

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

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, *92*(1), 117–125. https://doi.org/10.1037//0022-0663.92.1.117

Nilsson, N. C., Nordhal, R., & Serafin, S. (2016). Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, *12*(2), 108–134. https://doi.org/10.17011/ht/urn.201611174652

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General*, *129*, 242–261. https://doi.org/10.1037//0096-3445.129.2.242

Ott, M., & Freina, L. (2015). *A literature review on immersive virtual reality in education*: *state of the art and perspectives*. Proceedings of the eLearning and software for education Conference, Volume 1, Bucharest, Romania. pp. 133–141. https://doi.org/10.12753/2066-026X-15-020

Parong, J., & Mayer, R. E. (2018). Learning in immersive virtual reality. *Journal of Educational Psychology*, *110*(6), 785–797. https://doi.org/10.1037/edu0000241

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction*, *29*, 128–140. https://doi.org/10.1016/j.learninstruc.2013.02.006

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotions in cognitive load theory. *Educational Psychology Review*, *31*, 339–359. https://doi.org/10.1007/s10648-019-09473-5

Poythress, M. L., Russell, C., Siegel, S. I., Tremoulet, P. D., Craven, P., Berka, C., … Milham, L. (2006). Correlation between expected workload and EEG indices of cognitive workload and task engagement. In D. Schmorrow, K. Stanney, & L. Reeves (Eds.), *Augmented cognition: Past, present and future* (pp. 32–44). Arlington, VA: Strategic Analysis, Inc.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher educations: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Samsung Business USA. (2016). Is virtual reality ready for the classroom? https://image-us.samsung.com/SamsungUS/b2b/resource/2016/06/29/INFOGRAPHIC_VR_in_EDU_Survey_JUN16AP_1.pdf

Schmolck, H., Buffalo, E. A., & Squire, L. R. (2000). Memory distortions develop over time: Recollections of the O.J. Simpson trial verdict after 15 and 32 months. *Psychological Science*, *11*(1), 39–45. https://doi.org/10.1111/1467-9280.00212

Schrader, C., & Bastiaens, T. (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. https://doi.org/10.1016/j.chb.2011.11.011

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, *5*(258), 1–17. https://doi.org/10.3389/fpubh.2017.00258

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641–656. https://doi.org/10.1037/h0063404

Storbeck, J., & Clore, G. L. (2008). Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and Personality Psychology Compass*, *2*(5), 1824–1843. https://doi.org/10.1111/j.1751-9004.2008.00138.x

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. https://doi.org/10.1007/978-1-4419-8126-4

The Body VR LLC. (2016). The Body VR: Journey Inside a Cell [Computer Application]. Retrieved from https://thebodyvr.com/journey-inside-a-cell/

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*(1454), 1–22. https://doi.org/10.3389/fpsyg.2017.01454

Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology*, *104*(2), 485–498. https://doi.org/10.1037/a0026609

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

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

How to cite this article: Parong J, Mayer RE. Cognitive and affective processes for learning science in immersive virtual reality. *J Comput Assist Learn*. 2021;37:226–241. https://doi.org/10.1111/jcal.12482