**Title**
Cognitive and Affective Mechanisms of Immersive Virtual Reality Learning Environments

**Author**
Parong, Jocelyn

**Publication Date**
2019

UNIVERSITY OF CALIFORNIA

Santa Barbara

# Cognitive and Affective Mechanisms of Immersive Virtual Reality Learning Environments

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Psychological and Brain Sciences

by

Jocelyn Ann Natera Parong

Committee in charge:

Professor Richard E. Mayer, Chair

Professor Mary Hegarty

Professor Michael Miller

Professor Matthew Turk

December 2019

The dissertation of Jocelyn Ann Natera Parong is approved.

Mary Hegarty

Michael Miller

Matthew Turk

Richard E. Mayer, Committee Chair

November 2019

Cognitive and Affective Mechanisms of Immersive Virtual Reality Learning Environments

by

Jocelyn Ann Natera Parong

# ACKNOWLEDGEMENTS

This research was supported by the Office of Naval Research under Grant N000141612046.

I would like to extend my deepest gratitude to my advisor, Dr. Richard Mayer, who has provided invaluable guidance and feedback throughout my graduate career and has been a great mentor. I would also like to thank the members of my committee, Dr. Mary Hegarty, Dr. Michael Miller, and Dr. Matthew Turk, for their insightful comments and suggestions.

A special thanks to my wonderful research assistants who helped me during the data collection process: Britney Barrera, Thomas Crumly, Tal Horovitz, Angela Kempf-Luu, Anne Lei, Chloe Mneimne, Rena Rodriguez, and Julia Woods.

Thank you to my family and friends for their love and support, including my dad, Samuel Parong, and fiancé, Rogelio Gonzalez. Finally, this dissertation is dedicated to my mother, Marilou Parong, who has been my greatest supporter and always will be. Rest in peace. I did it, Mom!

VITA OF JOCELYN ANN NATERA PARONG
December 2019

## EDUCATION

**University of California, Santa Barbara,** Santa Barbara, CA — Dec. 2019
Ph.D. in Psychological and Brain Sciences
Dissertation: Cognitive and Affective Mechanisms of Immersive Virtual Reality Learning Environments

**California State University San Marcos**, San Marcos, CA — Dec. 2015
M.A. in Experimental Psychology
Thesis: Immediate Effects of Exergames on Executive Function

**Washington State University**, Pullman, WA — Dec. 2012
B.S. in Psychology, Magna Cum Laude

## RESEARCH EXPERIENCE

**Graduate Division Dissertation Fellow** — Sep. 2019 – Dec 2019
*University of California, Santa Barbara*

**Graduate Student Intern (paid position)** — July 2019 – Sep. 2019
U.S. Army CCDC Army Research Laboratory West, Playa Vista, CA
2019 Summer Student Symposium Oral Presenter Award
Mentor: Dr. Kimberly Pollard

**Graduate Student Researcher (funded position)**
Laboratory for Research on Learning and Instruction, *University of California, Santa Barbara* — Apr. 2016 – June 2019
Advisor: Dr. Richard Mayer

**Graduate Research Assistant**
Laboratory for Research on Learning and Instruction, *University of California, Santa Barbara* — Sep. 2015 – Mar 2016
Advisor: Dr. Richard Mayer
Cognitive Psychology Laboratory, *California State University San Marcos* — Aug. 2013 – Dec. 2015
Advisor: Dr. Dustin Calvillo

## TEACHING EXPERIENCE

**Teaching Associate**
Laboratory in Advanced Research Methods — June 2018 – Aug. 2018
*University of California Santa Barbara*

**Graduate Teaching Assistant**
Social Memory, *University of California, Santa Barbara* — Jan. 2016 – Mar. 2016
Instructor: Dr. Alex Swan
Educational Psychology, *University of California, Santa Barbara* — Sep. 2015 – Dec. 2015
Instructor: Dr. Richard Mayer
Introduction to Psychology [Section Instructor] — Aug. 2014 – May 2015

*California State University San Marcos*

Introductory Statistics in Psychology, *California State University San Marcos* Jan. 2014 – May 2015

Instructor: Dr. Marilyn Johnson-Kozlow

Research Methods in Psychology, *California State University San Marcos* Aug. 2014 – Dec. 2014

Instructor: Dr. Roger Morrissette

Laboratory in Cognitive Psychology, *California State University San Marcos* Aug. 2013 – July 2015

Instructors: Dr. Michael Diehr, Dr. Dustin Calvillo,
Dr. Carrick Williams, Dr. Evan Raiewski

## **PUBLICATIONS**

Parong, J., & Mayer, R. E. (under review). Cognitive and emotional distractions of learning in immersive virtual reality. *Contemporary Educational Psychology.*

Wells, A., Parong, J., & Mayer, R. E. (under review). Limits on training inhibitory control with a focused video game. *Contemporary Educational Psychology.*

Parong, J., Wells, A. & Mayer, R. E. (2019). Replicated evidence for a cognitive theory of game-based training. *Journal of Educational Psychology.* Advance online publication. https://doi.org/10.1037/edu0000413

Parong, J., & Mayer, R. E. (2019). Cognitive consequences of playing brain training games in immersive virtual reality. *Applied Cognitive Psychology, 2019,* 1-10. https://doi.org/10.1002/acp.3582

Mayer, R., E., Wells, A., Parong, J., Howarth, J. T. (2019). Learner control of the pacing of an online slideshow lesson: Does segmenting help? *Applied Cognitive Psychology, 33,* 930–935. https://doi.org/10.1002/acp.3560

Mayer, R. E., Parong, J., & Bainbridge, K. (2019). Young adults learning executive function skills by playing focused video games. *Cognitive Development, 49*, 43-50. https://doi.org/10.1016/j.cogdev.2018.11.002

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

Parong, J., Mayer, R. E., Homer, B., Plass, J, Macnamara, A. (2017). Learning executive function skills by playing focused video games. *Contemporary Educational Psychology, 51*, 141-151. https://doi.org/10.1016/j.cedpsych.2017.07.002

Calvillo, D. P., Parong, J. A., Peralta, B., Ocampo, D., & Van Gundy, R. (2016). Sleep increases susceptibility to the misinformation effect. *Applied Cognitive Psychology, 30*(6), 1061-1067. https://doi.org/10.1002/acp.3259

Calvillo, D., Parong, J. (2016). The misinformation effect is unrelated to the DRM effect with and without a DRM warning. *Memory, 24*(3), 324-333. https://doi.org/10.1080/09658211.2015.1005633

## **PRESENTATIONS**

Parong., J. & Mayer, R. (2019, November 14-17). *Is immersive virtual reality too distracting for learning?* [Poster presentation]. 60th Annual Meeting of the Psychonomic Society, Montreal, QC, Canada.

Wells, A., Parong, J., Mayer, R., MacNamara, A., Plass, J., & Homer, B. (2019, May 23-26). *Training inhibition with a focused video game* [Poster presentation]. 31st Annual Convention of the Association for Psychological Science, Washington D.C., United States.

Parong, J. & Mayer, R. E. (2018, November 15-18). *Cognitive and physiological mechanisms of virtual reality learning environments* [Poster presentation]. 59th Annual Meeting of the Psychonomic Society, New Orleans, LA, United States.

Parong, J. & Mayer, R. E. (2018, May 24-27). *Brain training in immersive virtual reality* [Poster presentation]. 30th Annual Convention of the Association for Psychological Science, San Francisco, CA, United States.

Parong, J. & Mayer, R. E. (2018, April 13-17). Learning science in immersive virtual reality [Paper presentation]. In Harley, J. M. (Chair), *Innovative Research on Virtual Reality, Augmented Reality, and Simulations*. 2018 Annual Meeting of the American Educational Research Association in New York City, NY, United States.

Parong, J., Mayer, R. E., Fiorella, L., MacNamera, A., Homer, B, & Plass, J. (2018, April 13-17). *Learning executive function skills by playing focused video games* [Poster presentation]. 2018 Annual Meeting of the American Educational Research Association in New York City, NY, United States.

Parong, J. & Mayer, R. E. (2017, November 9-12). *Learning science in immersive virtual reality* [Poster presentation]. 58th Annual Meeting of the Psychonomic Society, Vancouver, BC, Canada.

Mayer, R. E. & Parong, J. (2017, June 8-10). *Young adults learning executive function skills by playing focused video games* [Paper presentation]. In Homer, B. D. (Chair), *Using Video Games to Enhance Executive Functions Across the Lifespan.* 47th Annual Meeting of the Jean Piaget Society, San Francisco, CA, United States.

Parong, J. & Mayer, R. E. (2017, May 11). *Enhancing executive function through focused video games* [Paper presentation]. Annual Psychological and Brain Sciences Mini-Convention, University of California, Santa Barbara, CA, United States.

Parong, J., & Mayer, R. E. (2016, November 17-20). *Playing a focused game enhances executive function* [Poster presentation]. 57th Annual Meeting of the Psychonomic Society, Boston, MA, United States.

Calvillo, D. P., Parong, J. A., Peralta, B., Ocampo, D., & Van Gundy, R. (2015, November 19-22). *Sleep and retention interval increase the susceptibility to misinformation* [Poster presentation]. 56th Annual Meeting of the Psychonomic Society, Chicago, IL, United States.

Calvillo, D., Parong, J., & Mills, N. (2014, November 20-23). *The misinformation effect is unrelated to the DRM effect with and without a DRM warning* [Poster presentation]. 55th Annual Meeting of the Psychonomic Society, Long Beach, CA, United States.

Mills, N., Parong, J., & Calvillo, D. (April, 2015). *The effects of ego depletion on lineup rejection and the confidence-accuracy relationship*. Poster presented at the Annual Convention of the Western Psychological Association, Las Vegas, NV.

Mills, N., Hale, N., Parong, J., & Calvillo, D. P. (2014, May 22-25). *Forced rapid responding eliminates the own race bias in a facial recognition task* [Poster presentation]. 26th Annual Convention of the Association for Psychological Science, San Francisco, CA, United States.

Parong, J., Mills, N., & Calvillo, D. P. (2014, May 22-25). *Gender differences in the misinformation effect* [Poster presentation]. 26th Annual Convention of the Association for Psychological Science, San Francisco, CA, United States.

# ABSTRACT

## Cognitive and Affective Mechanisms of Immersive Virtual Reality Learning Environments

By

Jocelyn Ann Natera Parong

Students and educators value the potential use of immersive virtual reality (IVR) in the classroom to teach academic content as it may increase interest and motivation to learn, which in turn may increase learning outcomes. However, one criticism is that features of IVR, such as extraneous sounds, animations, and interactions, that are not relevant to the content of the lesson, as well as the affective arousal associated with the use of IVR, may be distracting to learning processes. To examine this distraction hypothesis, two experiments were conducted in which students viewed a biology (Experiment 1) or history (Experiment 2) lesson either in IVR or in a desktop lesson containing the same content, with or without practice questions. In both experiments, students who viewed the lesson on the desktop in a PowerPoint (Experiment 1) or an interactive video (Experiment 2) outperformed those who viewed the IVR lessons on transfer tests. The desktop lessons led to higher cognitive engagement based on EEG measures in both experiments, and less self-reported extraneous cognitive load in Experiment 1. Participants also reported more high-arousal positive emotions after the IVR lessons in both experiments, and experienced higher physiological arousal based on heart-rate measures after the IVR lessons compared to the desktop lessons in Experiment 2. The same pattern of results was found when adjunct practice questions were or were not included in the lessons. Across both experiments, mediation analyses suggest that the negative relationship between instructional media and learning outcomes can be explained by self-reported cognitive processing or emotional arousal, particularly for retention test performance. Overall, immersive environments may create high emotional arousal and cognitive distraction during learning, which leads to poorer learning outcomes than desktop environments.

TABLE OF CONTENTS

# Chapter I: Introduction

## Objective and Rationale

Immersive virtual reality (IVR) may be defined as a technology that simulates sensations (e.g., visual, auditory, or haptic experiences) and behaviors (e.g., flying, walking, being in a location, etc.) in a virtual environment (LaValle, 2019). As IVR consoles, such as the HTC Vive, Oculus Rift, or Samsung Gear, become more accessible in classrooms, it is important to examine and understand their role in academic learning. Are these new media more effective than more traditional instructional methods, such as a lecture and PowerPoint? Proponents of using IVR to display academic lessons argue that the novel immersive and interactive features of this technology may have motivational factors to pique student interest in learning and increase learning outcomes (Bailenson & Blascovich 2011). Indeed, many studies and consumer surveys report both high student and teacher support for using IVR in classrooms; for example, a Samsung consumer survey reported that 83 percent of teachers believe IVR will improve learning outcomes (Hew & Cheung, 2010; Makransky et al., 2019; Parong & Mayer, 2018; Samsung Business USA, 2016). Concurrent with the promise of IVR revolutionizing education, IVR offers many affordances over traditional technologies, including the use of embodied pedagogical agents, enhanced visualizations, and motivational factors associated with immersion and presence (Bailenson et al., 2008; Bailenson & Blascovich, 2011).

However, one drawback of using this new medium may be related to its effects on cognitive and/or affective processing during the lesson. Although learners may be more interested and feel more involved in learning from IVR, certain features of the VR environment may be distracting to learning processes, such as disrupting cognitive processing during information acquisition or creating excessive emotional arousal that detracts from learning, which can be explained by the cognitive theory of multimedia learning and cognitive load theory. Because of this, another important question is to determine whether asking students to engage in a generative learning strategy, such as answering practice questions at various points throughout the lesson, could improve learning outcomes from IVR. The goals of this study were to examine these issues. This study aimed to determine (1) whether IVR is an effective educational medium compared to more traditional instructional methods, such as a desktop lesson, (2) whether learning outcomes from these media can be explained by cognitive or affective processes, and (3) whether these multimedia lessons can be improved by adding a generative learning strategy, particularly without diminishing enjoyment of the lesson.

## Affordances of IVR: The Motivator Hypothesis

The affordances of using IVR to display academic lessons include practical and theoretical affordances related to student motivation and emotional arousal. The main practical affordance of IVR is that it has opened up new possibilities for delivering academic information to students. IVR can simulate any location with any number of customizable pedagogical agents or co-learners. It can be used to deliver lessons that could potentially be dangerous, expensive, or otherwise unfeasible in the real world. For example, learners can gain access to previously inaccessible environments, such as traveling to space or entering an animal cell, or practice dangerous or expensive tasks, such as a chemistry experiment, surgical procedure, or train in a military battlefield. These practical affordances are the main motivation that teachers report for using IVR in classrooms (Ott & Freina, 2015).

Along with the practical affordances, some of the underlying theoretical affordances of using IVR over other media revolve around student motivation. Instructional media could influence learning outcomes by increasing motivation to learn and boosting cognitive processes that are associated with motivation, such as attention and focus. Two pathways that could increase motivation and cognitive processing during learning are the feeling of presence during a lesson and emotions experienced during a lesson.

### ***Presence***

First, unlike other instructional methods, IVR is an immersive experience which causes a learner to feel a sense of presence. It is customary to distinguish between the two closely related constructs of immersion and presence. Although both can be generally described as involving being or feeling surrounded by something, immersion can be defined as more objective physical, perceptual, or cognitive demands of the interaction, whereas presence can be defined as the psychological subjective feeling of "being there" (Nilsson et al., 2016). A few traditional views of immersion have been defined: as physical properties of the system, as a perceptual response to the system, or as a cognitive response to a narrative or to challenges. Immersion as a property of the system can be described as the physical properties and potential output of the system itself (e.g., the HTC Vive). The higher the fidelity of the displays and tracking, the greater the level of immersion (Slater, 2003). Immersion as a perceptual response describes the sensory representations caused by the system; the higher the number of sensory modalities engaged (e.g., sight, sound, smell), the higher the level of immersion (McMahan, 2003). Immersion as a response to narratives is the sensation of being mentally absorbed by a storyline, characters, or fictional world; the more compelling the narrative, the more immersive the media (Ryan, 2001, 2008). Immersion as a response to challenges describes one's engagement with the challenges within the media. For example, in games it may be brought on by the desire to gain points or by devising a specific strategy (McMahan, 2003). Similarly, in an academic lesson it may be induced by a desired level of difficulty for understanding the material.

Although these facets of immersion could be incorporated into a lesson in IVR, immersion may not be completely unique to IVR. For example, a narrative in a fictional book or movie can also be immersive on some dimensions. On the other hand, presence, may be unique to IVR in that it allows learners to feel more psychologically present in a virtual simulation (compared to being present in the real world) than is possible in traditional classrooms, even with the use of computers (Kafai, 2006).

Presence in an immersive environment is experienced by a two-step process (Wirth et al., 2007). First, the user uses spatial cues to perceive that the virtual environment is a plausible space. Second, the user then experiences him or herself being located in the space with perceived possibilities to act. The virtual environment is more likely to be perceived as a plausible space that can be acted upon if visual and auditory cues are rich in quality and logically consistent with the real world. For example, some factors that may play a role in perceived presence within these two steps include (1) control factors, such as degree and immediacy of control, (2) sensory factors, such as environmental richness and multimodal presentation, (3) distraction factors, such as isolation and interference awareness, and (4) realism factors, such as information consistency with the objective world (Witmer & Singer, 1998). Thus, from this rationale, 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, should elicit greater psychological presence. A meta-analysis reports that across different levels of immersive technologies, immersion had a medium-sized effect on presence (Cummings & Bailenson, 2016).

Feeling present in an immersive environment may increase motivation to be engaged and attentive in the simulation. For example, Yee and Bailenson (2006) found that students who took the view of a senior citizen in an IVR environment developed more empathy toward elderly adults than students who took the view of a young person. In turn, learners would be more motivated to persist in a lesson and, in turn, increase outcomes from the goals of the lesson. Previous research has shown that presence is associated with learning outcomes. (e.g., Makransky & Lilliholt, 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 associated with performance on transfer tests.

### ***Emotional Arousal as an Enhancer***

Additionally, the novelty of using and interacting with a new technology or the immersive instructional design may induce positive emotional arousal during a lesson. Pekrun (2014, 2017) describes four groups of emotions in the context of academic learning. Achievement emotions relate to success and failures resulting from activities (e.g., pride related to success or shame related to failure on an exam). Epistemic emotions are caused by cognitions about the learning task (e.g., surprise about a new task, confusion about an obstacle, or delight when a problem is solved). Topic emotions are caused by the content in the lesson (e.g., empathetic emotions towards the characters, or delight or displeasure based on one’s interest in the subject). Social emotions involve relationships between students’ teachers and peers in the classroom. Of particular interest in this study are the effects of a lesson on topic emotions, which influences a learner’s immediate attention and motivation during a lesson, and in turn learning outcomes from the lesson. It is posited that IVR could increase topic emotions. For example, presence could increase emotional investment in characters in a lesson. Enjoyment of using a novel technology could also increase pleasure in the topic presented.

Emotions are largely interconnected with cognitive processes that occur during learning; for example, emotions have been shown to be involved in multiple cognitive mechanisms that play a large role in learning, including memory, attention, and perception (Izard, 2009; Lewis, 2005; Tucker, 2007). The Cognitive-Affective Theory of Learning with Media (CATLM) proposes motivation to learn is mediated by a person’s affect as it increases or decreases cognitive engagement (Moreno & Mayer, 2007). Similarly, the Integrated Cognitive Affective Model of Learning with Multimedia (ICALM) proposed by Plass and Kaplan (2015) describes how the design of the multimedia environment affects both the learner’s emotions and cognitive processes. They propose that just as instructional design should evoke appropriate cognitive processing, it should also evoke efficient emotional experience; instructional design should be compatible with the learner’s emotional self-regulation skills and should minimize emotional overload.

Extant research on emotions and memory has shown that positive emotions facilitates encoding during learning and helps retrieval of information (see Storbeck & Clore, 2008; Tyng et al., 2017 for reviews). Specifically, previous research has shown that the emotional design of multimedia learning material can induce positive emotions in learners that in turn facilitate comprehension and transfer. For example, in a lesson about immunization, using characters with round faces and warm colors induced positive emotions in learners, which in turn improved learning outcomes from the lesson compared to characters with square faces with no color (Plass et al., 2014; Um et al., 2012).

In another study, Mayer and Estrella (2014) found that students who viewed a biology lesson with enhanced graphics using colored characters with expressive faces had better learning outcomes and reported higher effort than students who viewed a lesson with black-and-white characters with non-expressive faces. As students generally report more positive emotions during an IVR lesson compared to other media (e.g., Parong & Mayer, 2018), one could predict that IVR should also lead to enhanced comprehension and transfer. Positive emotional arousal during a lesson can increase motivation to engage in the lesson, to exert more effort in understanding, and to be resilient when faced with obstacles in learning (Mayer, 2009; Wentzel & Miele, 2016).

### *Motivational Theories*

The links between emotions, motivation, and learning outcomes may be explained by interest theory, self-determination theory, and self-efficacy theory. Interest theory states that students exert more effort in learning when they value the material. Interest can be described as an individual's psychological state that is characterized by increased attention, effort, and affect caused by a particular moment (situational interest) or a predisposition to reengage with a certain topic over time (individual interest; Harackiewicz et al., 2016; Renninger & Hidi, 2016; Schiefele, 2009). Research has shown that students' self-rated individual interest in a subject has been associated with school performance as measured by grades and test scores (Schiefele et al., 1992). Situational interest can turn into individual interest for a subject over time. First, features of the environment (e.g., novelty with a new technology) induce situational interest. Over time, repeated experience of this situational interest can spark an individual's desire to reengage with the material, which then develops into a self-sustaining, well-developed, individual interest that can be seen in other contexts (Renninger & Hidi, 2016). However, research on seductive details intended to prime situational interest shows that adding interesting, but irrelevant, material to a multimedia lesson may hurt learning (Harp & Mayer, 1997, 1998; Mayer et al., 2008).

During a lesson, some positive emotions, such as enjoyment, can increase both a learner’s individual and situational interest. On the other hand, positive emotions like relaxation, can decrease a learner’s situational interest and reduce motivation to continue making an effort in learning. Negative emotions, such as anxiety, tend to decrease both situational and individual interest for learning, but can also induce individual interest to invest more effort, particularly if the student expects to succeed (Pekrun, 2014). Viewing a lesson in various media may induce different levels of interest, and in turn emotions. This study aims to explore the differences in emotion induced by the learner’s interest in desktop lessons versus IVR lessons. For example, if learners feel more enjoyment during an IVR lesson than a lesson using other media, this may increase interest in the lesson, and in turn, learning outcomes.

Self-determination theory states that there are three psychological needs: autonomy, competence, and relatedness. When these three components are fulfilled, positive emotions and intrinsic motivation during a task increase (Przybylski et al., 2010; Ryan & Deci, 2000). Autonomy is the feeling of control over one’s actions. Provisions for choice, informational feedback, and non-controlling instructions have all been shown to enhance autonomy and can induce positive emotions during a lesson. However, events or conditions that diminish a sense of choice, control, or freedom for either the means or ends of an action can interfere with perceived autonomy and undermine intrinsic motivation (Ryan et al., 2006). Competence is the need for challenge and feelings of effectance (Deci, 1971). Lessons that allow opportunities to acquire new skills or abilities or to be optimally challenged increase feelings of competence, and in turn increase intrinsic motivation, whereas contexts that are repetitive and require the same skill or are too easy or hard undermine feelings of competence, and in turn diminish intrinsic motivation (Ryan et al., 2006). Relatedness is experienced when one feels connected with others. Motivation to learn increases when the learner feels connected to the teacher or other learners. Instructional media may prime different levels of learners' self-determination and motivation. For example, learners may feel more of a sense of autonomy in an IVR lesson than a desktop lesson as the user has multiples sources of interaction (e.g., body and head movements, pressing a button) in IVR compared to only one source of interaction by pointing and clicking a computer mouse on a desktop. Thus, motivation to learn, and in turn learning outcomes, may be increased by an increased sense of self-determination.

According to self-efficacy theory, students exert more effort when they see themselves as competent for the task (Schunk & DiBenedetto, 2016). Self-efficacy is a person's judgments of his or her ability to perform a given task, and it affects the amount of effort and persistence a person exerts to perform that task (Bandura, 1977; Schunk, 1991). Research has shown that self-efficacy is related to engagement in a lesson as well as academic performance (Chemers et al., 2001; Pietsch et al., 2003; Schunk & Hanson, 1985). Within a lesson, interactions with a lesson or instructor may induce feelings of self-efficacy (e.g., feedback from the instructor or being able to advance to the next section in a digital lesson). Additionally, an appropriate level of challenge for understanding the material may prime feelings of self-efficacy as students may feel competent in following along with the lesson. Lessons in different instructional media may induce different levels of self-efficacy and emotions. Users receive feedback from multiple types of interactions in IVR (e.g., both selection interactions and physical interactions), whereas they may only receive feedback after a button press in a desktop lesson. Therefore, IVR users may feel a higher sense of competence and self-efficacy than those who view desktop lessons, and this competence could translate into higher learning outcomes. In sum, the motivator hypothesis of IVR suggests that IVR increases learning outcomes by increasing motivation and cognitive processing through presence and emotions, which are supported by motivational theories.

## Drawbacks of IVR: The Distractor Hypothesis

### *Cognitive Theory of Multimedia Learning*

In contrast to serving as a motivator, using IVR to display academic lessons may serve as a distractor during learning. The Cognitive Theory of Multimedia Learning (CTML) proposes that lesson design (i.e., the way in which pictures and words are presented during a lesson) can influence how incoming information is processed, and in turn affect learning outcomes (Mayer, 2009, 2014). CTML, which was derived from cognitive load theory (Paas & Sweller, 2012, 2014; Sweller et al., 2011), outlines principles for designing elements within a multimedia lesson (i.e., learning from words and pictures) congruent with cognitive theories of how humans actively process information (Mayer, 2009). First, CTML describes the select-organize-integrate model (SOI), which explains how learners make meaning of incoming information during learning, which corresponds to Atkinson and Shiffrin's (1968) multi-store model of memory. First, leaners *select* the most relevant incoming information (e.g., pictures or words) in sensory memory. Then, learners must build relevant connections based on the information's underlying structure to *organize* the selected information into a coherent mental representation in working memory. Finally, learners *integrate* the new mental representation from working memory with existing knowledge structures stored in long-term memory, such as schemas, categories, or principles (Mayer, 2009, 2014).

As the learner engages the SOI model during a lesson, he or she experiences three types of cognitive processing, which affects how effective the SOI processes occur. The three types of cognitive load during learning include extraneous processing, which does not serve an instructional goal and is due to poor instructional design or poor selection of material; essential processing, which is the processing necessary for holding and organizing relevant information in working memory; and generative processing, which is needed for deeper understanding of the material and is imperative in organizing and integrating information. A learner's limited cognitive capacity is divided among these three cognitive processing, and effective learning occurs when cognitive resources are more focused on essential and generative processing, and less focused on extraneous processing. Based on this model, a set of evidence-based principles were derived to help students reduce extraneous processing, manage essential processing, and foster generative processing (Fiorella & Mayer, 2015; Mayer, 2009; Mayer & Fiorella, 2014; Mayer & Pilegard, 2014).

In IVR, lessons may inherently draw some of the learner's cognitive resources to extraneous processing. For example, there may be animations, sounds, or interactions with the lesson that are superfluous to the main content of the lesson. One principle in CTML that simulation in IVR tend to violate is the coherence principle, which states that multimedia lessons should eliminate extraneous words, pictures, animations, etc. that are not essential to the lesson. Adding additional material can hinder the learner's cognitive processing of incoming information by interfering with the selection process (e.g., taking the learner's attention away from relevant material), the organization process (e.g., inserting extraneous material into the coherent mental model), or the integration process (e.g., priming irrelevant prior knowledge; Mayer, 2002). For example, the IVR biology simulation used in Parong and Mayer (2018), *The Body VR: Journey Inside a Cell*, immersed the learner in 360-degree views of the inside of a blood vessel and cell. During the lesson, the learner was constantly exposed to moving parts surrounding him or her (e.g., interactive red blood cells moving past the learner), while tasked with listening to the narrator explain the parts and functions of the blood vessel cells. The coherence principle would predict that learners in IVR would perform worse on transfer tests compared to other media that better adhere to the coherence principle, such as a PowerPoint slideshow with only essential words and pictures, which is what the researchers found from the students' learning outcomes. 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). Additionally, research on seductive details intended to prime situational interest shows that adding interesting but irrelevant material to a multimedia lesson can hurt learning (Harp & Mayer, 1997, 1998; Mayer et al., 2008). This additional extraneous load may reduce cognitive resources to devote to essential and generative load and cause poorer learning outcomes.

### ***Emotional Arousal as a Distractor***

Although the previous section highlighted potential benefits of emotions during learning, emotions can also be detrimental to learning. Based on Russell's (2003) description of core affect, emotions lie on two dimensions, pleasure and activation. Pleasure, or valance, of the emotion can be positive or negative; activation, or arousal, can be low or high. High arousal events (both negative and positive valence) are often reported to be remembered more vividly than neutral events (e.g., Canli et al., 2002; Charles et al., 2003; Ochsner, 2000), but are also less accurate than believed to be (Schmolck et al., 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).

The idea of emotional arousal as a distractor to learning can be extended back to CTML and CTL; excessive emotional arousal may be thought of as a source of extraneous cognitive load, as it diverts cognitive resources to processing the emotion-evoking stimuli or the emotions themselves. Both positive and negative high arousal emotions during learning, such as being extremely excited or stressed, could induce task-extra processing, in which cognitive processes are devoted to emotional regulation that is not relevant to learning. Additionally, high arousal emotions could cause task-irrelevant processing in which highly arousing emotions lead to the retrieval of irrelevant information, such as seductive details, not essential to learning (Plass & Kalyuga, 2019). In sum, the distractor hypothesis of IVR suggests that IVR decreases learning outcomes by disrupting learning processes and increasing extraneous cognitive processing through cognitive or affective distractions.

## Media Comparison Evidence

Media comparison studies have traditionally compared learning outcomes between two media. For example, one group of learners could view a lesson using a conventional method of delivering academic information, such as a text passage, and another group of learners could view a lesson in a newer mode of delivery, such as a simulation game displayed on a desktop computer. The two groups could then be compared in their learning outcomes, which are typically measured by a retention or transfer test of the learned material (Mayer, 2014). Retention tests measure how well the learners remember the material from the lesson, whereas transfer tests measure how well the learners apply what they learned to a new task or situation (Mayer & Wittrock, 1996).

A meta-analysis of media comparisons between lessons in computer games compared to conventional media with the same content found two academic domains, science and second-language learning, in which students learned more from the game than a conventional media (Mayer, 2014). Additionally, for simulation games, similar to immersive virtual reality simulations that are used in this study, there was an overall medium to large effect size ($d = 0.62$) favoring games over conventional media (Mayer, 2014).

### ***Criticisms of Media Comparison Studies***

It should be noted that there are criticisms of media comparison studies, which argue that the media themselves do not cause learning, but rather what has more value in being studied is the instructional method (Clark, 1983). Clark (1994) argued that different media can accomplish the same learning goals, and no single attribute of a medium can contribute to a unique cognitive effect on learning. The methodological issue with media comparison studies is that they may confound the medium and the instructional method as they typically both vary between two media. For example, IVR lessons using a head-mounted display technology may implement higher definition visualizations in the instructional lesson compared to lower definition visualizations in a lesson on a desktop computer.

However, Kozma (1994) argued that there is merit in comparing different media as there is still a need to determine the relationship between media and learning and explain what causes that relationship. Certain media may have specific characteristics that affect the ways in which learners represent relevant characteristics and make them more or less suitable for learning, including the physical, mechanical, or electronic characteristics of the technology itself, or the way in which the pictures, animations, text, or spoken language are presented to the learner based on the processing capabilities of the technology (Salomon, 1979). From this perspective, underlying theories about the way humans process incoming information, such as the cognitive load theory, can help explain how particular features in multimedia and technologies affect learning, and therefore add value in media comparison studies.

### ***IVR and Transfer of Learning***

As the goal of education is for students to transfer what they learned to new situations and solve problems that they were not explicitly taught to solve, it is important to examine learning outcomes of various instructional methods (Mayer & Wittrock, 1996). However, there has been little scientific research on the use of IVR in education in the past decade. Meta-analyses have reported that many studies rely on non-experimental methods or descriptive analyses based on student surveys and interviews to determine the effectiveness of virtual reality technologies in the classroom (Hew & Cheung, 2010; Ligorio & van Veen, 2006). For example, one study examined learning outcomes from a computer graphics class in which students used an immersive 3D program to learn about function-based shape modeling. The researchers used exam scores throughout the course to determine the efficacy of IVR and reported a 14 percent increase after one semester (Sourin et al., 2006). However, the researchers did not include a control group in their experiment, so the increase in exam scores cannot be attributed to solely the use of IVR as it may have been due to other factors within the classroom or students.

While some researchers have found effects favoring IVR over desktop lessons (Kozhevnikov et al., 2013; Webster, 2016), others have found that IVR hinders learning (Makransky et al., 2019; Moreno & Mayer, 2004; Parong & Mayer, 2018). Kozhevnikov and colleagues (2013) presented students with a lesson on relative motion concepts in either an immersive virtual environment using a head-mounted display (IVE) or on a desktop virtual environment presented on a computer (DVE). After the lesson, the IVE group performed significantly better than the DVE group on solving two-dimensional relative motion problems, suggesting a further transfer of learning from the use of immersive virtual reality compared to the use of virtual reality on a computer. In another example, military participants learned about basic corrosion prevention and control either in an IVR environment using a combination of a head mounted display from WorldViz and movement tracking with the Microsoft Kinect or a desktop multimedia presentation. Those in the IVR environment had significantly higher gains in learning than those in desktop learning environment (Webster, 2016).

Contrary to these findings, Parong and Mayer (Experiment 1, 2018) found that viewing a cell biology lesson in IVR led to worse learning outcomes than viewing an equivalent lesson in a self-paced PowerPoint lesson. However, this is consistent with the coherence principle of the cognitive theory of multimedia learning as the IVR lesson did not eliminate extraneous animations and music in its lesson. This could have caused extraneous cognitive load in the learner, leaving fewer resources for essential cognitive load required to process the important material, and therefore leading to worse learning outcomes. Moreno and Mayer (2004) found similar results when comparing a plant biology lesson using an immersive head-mounted display to a less immersive desktop lesson. Although the students reported higher presence using the head-mounted display, they did not perform significantly better on tests of transfer or retention. Because of these mixed findings, it is important to further investigate which factors within IVR influence learning outcomes.

## Generative Learning Strategies to Improve IVR

As students generally report liking IVR lessons, there may be some merit in adding generative learning strategies to boost performance, rather than attempting to redesign lessons and possibly detract from their entertainment value. Extant research in generative learning strategies added to lessons in various types of media has shown that they can increase learning outcomes. The third goal of this study is to determine whether the effects of a specific generative learning strategy, practice testing, can be extended to lessons in IVR.

### *Generative Learning Theory*

According to generative learning theory, meaningful learning occurs when learners engage in appropriate cognitive processing during learning (i.e., appropriately selecting, organizing, and integrating incoming information with previous knowledge; Fiorella & Mayer, 2015). In particular, the organizing and integrating steps of the SOI model involve generative processing, which is defined by Wittrock (1989) as making mental representations and building connections between different elements of the verbal and pictorial models of the just-learned information and the learner's existing knowledge. One strategy to foster generative processing is through learner-based activities (e.g., adding worksheets to a lecture or asking the student to self-explain during the lesson) through which the learner can be guided in building and forming those connections. Previous research has shown that some generative learning strategies have stronger effects on learning, including summarizing and practice testing, while other generative learning strategies are less effective, such as rereading and highlighting (see Dunlosky et al., 2013 and Fiorella & Mayer, 2015 for reviews).

### *Practice Testing*

One example of an effective generative learning strategy is practice testing, which is defined as a low-stakes or no-stakes practice activity; some examples include practicing recall with flashcards, completing practice problems at the end of a textbook chapter, or completing practice tests similar to the final test (Dunlosky et al., 2013). Practice testing triggers both direct and mediated effects on learning. The act of taking the test itself engages the elaborative retrieval process. It can also mediate the amount or kind of restudying after the test, which influences the type of encoding of the material. Over 100 years of research has repeatedly shown the robust effects of practice testing on improving retention and transfer (e.g., Roediger & Karpicke, 2006; Spitzer, 1939; see Roediger & Butler, 2011 for review). Dunlosky and colleagues (2013) proposed that practice testing is one of the highest utility learning strategies because effects have been seen across a number of test formats, domains, learner ages, outcome measures, and retention intervals. This study aims to extend the practice testing effect to a new medium, IVR.

## Current Study and Predictions

Over 2 experiments, the current study examined 3 research questions: (1) Are academic lessons in IVR as effective as an equivalent lesson displayed on a desktop? (2) What are the cognitive and affective mechanisms that can explain the learning outcomes from these media? (3) Does adding a generative learning strategy of practice testing during the lesson boost learning outcomes? The first experiment aimed to replicate and extend previous findings that a biology lesson is more effective when presented in a PowerPoint format than in IVR (Parong & Mayer, 2018), and to determine the underlying cognitive and affective mechanisms associated with learning. The second experiment was a novel extension of these findings to a sparsely studied academic domain, history, as research has tended to focus on STEM domains or technical training; it compared learning in IVR to learning from an interactive desktop video.

Based on CTML and previous work in our lab, the *distraction hypothesis* was tested in both experiments, which proposes that IVR causes the learner to waste limited processing capacity on cognitive or affective processing that does not support deep learning of the material, as compared to more conventional media. Specifically, our predictions were derived from the idea that IVR causes more extraneous cognitive processing and more highly arousing affective responses (which could also be a source of extraneous cognitive processing) during learning than a desktop lesson (i.e., PowerPoint or video), which distracts from learning and causes poorer learning outcomes. Both experiments also tested the practice testing effect based on generative learning theory in the form of oral questions asked throughout the lesson with both IVR and desktop media.

### *Predictions*

The predictions derived from the distraction hypothesis were as follows:

**Prediction 1**: Learners will perform better on retention and transfer tests after learning from a desktop lesson compared to learning from IVR.

**Prediction 2:** Cognitive processing between the IVR and desktop lessons will differ; learners will show higher levels of cognitive distraction during the IVR lessons than the desktop lessons.

**2a**: Based on a self-report survey, it was predicted that those who viewed the IVR lessons would report higher ratings of extraneous cognitive load and lower ratings of essential and generative cognitive load during learning than those who viewed the desktop lessons.

Additionally, electroencephalography-based (EEG) measures of cognitive processing during the lesson were included. EEG measures electrical activity produced by the brain in response to changing levels of cognitive stimuli via electrodes placed on the scalp (Anderson & Bratman, 2008). The continuous EEG signal is composed of oscillations in various frequencies, which reflect the cognitive processes involved in representing, maintaining, and transferring information within and across neuronal paths (Antonenko et al., 2010; Klimesch et al., 2005). Previous research has shown that changes across this signal can serve as direct and measurable indices of specific brain activities, which can then be correlated with cognitive processes related to attention, learning, and memory (Basar, 1999). Advanced Brain Monitoring's B-Alert x10 mobile EEG system provides proprietary measures of *cognitive engagement*, which is defined as a 4-class continuum of attention and focus on a task, including sleep onset, distraction, low engagement, and high engagement, and *cognitive workload*, which is defined as the amount of mental load a user experiences during a task, which includes 3 distinct classes of boredom, ideal workload, and cognitive overload. These have been shown to be valid indices of cognitive processing (Berka et al., 2004, 2007, Poythress et al., 2006).

**2b:** It was predicted that students in the IVR lessons would exhibit lower probabilities of being in a high cognitive engagement state and fewer participants classified to have ideal cognitive workload (compared to boredom or cognitive overload) during learning than students in the desktop lessons.

**Prediction 3:** Affective processing during learning will differ between the two media, and the IVR lessons will lead to higher levels of positive emotion with high arousal than the desktop lessons.

**3a:** It was predicted that learners in IVR would report higher levels of positive/high arousal emotions, such as happy or excited, and lower levels of negative/low arousal emotions, such as sad or bored, than the desktop groups based on a self-report survey.

As emotional arousal can be associated with physiological measures, measures of electrodermal activity (EDA), heart rate, and heart rate variability (HRV) were included. The input signals from both EDA and electrocardiography (ECG) used to calculate heart rate and HRV are presumed to reflect activity of the autonomic nervous system (ANS), which plays a major role in maintaining the internal equilibrium of the body between the antagonistic control of its two subcomponents, the sympathetic and parasympathetic systems. It has been shown that emotional stimuli impact ANS activity (Andreassi, 2010). EDA measures the amount of sweat produced on a person's skin, which is secreted by the eccrine sweat glands in response to sympathetic nervous activity (Dawson et al., 2007). EDA measures, such as 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, 2010; Bradley & Lang, 2007; Lang et al., 1993). The ECG signal receives input from the sino-atrial node in the wall of the right atrium of the heart, which acts as a pacemaker of cardiovascular activity, receives input from a post-ganglionic fiber from the sympathetic nervous system and a vagal nerve from the parasympathetic nervous system. Therefore, heart rate and changes in heart rate (i.e., HRV) are dependent on activity level from the ANS, and in turn emotional stimuli and regulated emotional responses (Appelhans & Luecken, 2006; Kim et al., 2004). Lower heart rate variability, specifically in the high frequency band, indicates higher arousal and has been found to be correlated with stress, panic, anxiety, and worry (Kim et al., 2018; Shaffer & Ginsberg, 2017).

**3b:** It was predicted that IVR groups would exhibit higher indications of physiological arousal according to EDA and heart rate measures than desktop groups.

Also of interest in this study were the links among instructional media, cognitive processes, affective processes, and learning outcomes, which were examined using forms of regression and mediation analyses.

**Prediction 4:** A proposed cognitive-affective distraction model of immersive virtual reality learning environments is shown in Figure 1a. 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 heart rate variability were used as measures of affective processing, and self-reported extraneous cognitive load and probability of high cognitive engagement were used as measures cognitive processing, as shown in Figure 1b. As immersion increases (comparing desktop to IVR), extraneous cognitive processing and extraneous affective processing will increase, which will cause a decrease in learning (both retention and transfer).

A secondary aim of this study was to investigate whether a generative learning strategy, practice testing, could be extended to a relatively novel medium, IVR, in two academic domains. Generative learning theory predicts that adding practice testing to an academic lesson should boost learning outcomes as it recruits cognitive resources to active processing of the just-learned material.

**Prediction 5:** Learners who received practice questions during the lesson will perform better on retention and transfer tests than students who did not receive practice questions.

**Prediction 6:** Practice testing will lead to less cognitive distractions and more appropriate processing than no practice testing.

**6a:** It was predicted that engaging in effective practice testing would result in higher ratings of essential and generative cognitive load, as well as lower ratings of extraneous cognitive load, than not engaging in practice testing.

**6a:** Similarly, students in the practice testing conditions were predicted to exhibit more learners with ideal cognitive workload and higher probabilities of being in a high engagement cognitive state during learning than students in the no practice testing conditions.

Finally, a goal of using generative learning strategies may be to not detract from the enjoyment or interest in the lesson (or optimistically boost enjoyment and interest). Previous research using the same biology lesson showed that there were no significant differences in self-reported emotions (both positive and negative) between an IVR lesson and an IVR lesson with a summarizing prompt included (Parong & Mayer, 2018).

**Prediction 7:** Practice testing will not cause any differences in affective processing based on valence or arousal compared to no practice testing.

**7a:** It was predicted that self-report measures of emotions during the lesson would not differ between practice testing and no practice testing groups.

**7b:** Finally, it was predicted that EDA and heart rate measures would not differ between practice and no practice testing groups.

In short, the generative learning strategy hypothesis predicts adding a requirement to engage in practice testing after each of the six sections of the lesson will result in higher post-test performance (learning outcome), focus cognitive processing during learning toward essential and generative processing rather than extraneous processing (cognitive processes), and not affect affective processes.

**Alternative prediction 7:** It is possible, however, that being required to give oral answers to the experimenter throughout the experiment may be stressful for learners, and thereby create excessive emotional distraction or physiological arousal.

# Chapter II: Experiment 1

The goal of Experiment 1 was to determine whether IVR or a PowerPoint was a more effective medium for delivering a biology lesson and to examine the cognitive and affective mechanisms associated with learning. Additionally, the effectiveness of adding a generative learning strategy to the lesson was examined.

## Method

### *Participants and Design*

Sixty-one participants (40 female, 20 male, 1 other) aged 18 through 38 years ($M$ = 20.51, $SD$ = 3.27) were recruited through the Psychology Paid Subject Pool through the Sona website at the University of California, Santa Barbara. Participants were compensated 20 dollars for completion of the experiment. A power analysis was conducted using G*Power 3.1 (Faul et al., 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) x 2 (practice testing vs. no practice testing) between-subjects design, 15 participants each served in the IVR-practice testing (IVRp) group, IVR-no practice testing (IVRn) group, and PPT-practice testing (PPTp) group, and 16 served in the PPT-no practice testing (PPTn) group.

### *Materials*

The computer-based materials consisted of two instructional lessons about the human blood stream: an immersive virtual reality simulation (IVR lesson) and a desktop slideshow (PPT lesson). The IVR lesson, called *The Body VR: Journey into a Cell*, consisted of a simulated trip through the human blood stream, which is shown in Figure 4 (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 lesson lasted approximately 12 minutes and ran either continuously (for the IVRn group) or was displayed in 6 segments of approximately 2 minutes each (for the IVRp group). In the IVRp version, the experimenter posed an orally-stated question and waited for an orally-stated answer from the participant after each of the 6 segments, which are listed in Appendix F. 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 5, which contained screenshots from the IVR lesson and corresponding printed text transcribed from the narration in the IVR lesson. The PPT slideshow displayed using DirectRT; it was self-paced, and participants were given unlimited time to view each slide, but could not return to previous slides. Participants spent an average of 8 minutes to complete the PowerPoint. The slideshow ran continuously (for the PPTn group) or was displayed in 6 segments of 2 to 7 slides each (for the PPTp group). In the PPTp version, there was an orally-stated question after each of the 6 segments, identical to the IVRp group.

The paper-based materials consisted of a pre-lesson questionnaire, post-lesson questionnaire, and post-test. The pre-lesson questionnaire (shown in Appendix A) solicited demographic information (i.e., the participant’s 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), activity checklist (checking which of 10 science-related statements applied to them, such as "I have participated in science fairs" or "I enjoy watching science documentaries"), and rating scale (rating their knowledge of the human body from 1 for "very low" to 5 for "very high”). The background academic survey, activity checklist, and rating scale were used to create a learner’s prior knowledge score by adding the number of science-related classes the learner took in high school and college, the number of items checked on the activity checklist, and a score between 0 and 3 on the learner’s rating of knowledge of the human body. The prior knowledge score was used to verify that groups did not differ in prior knowledge as it strongly 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 (Appendix C) asked participants to rate a series of 36 statements on a 7-point scale. Subscales included presence, which were 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* = 0.77), difficulty (2 items; e.g., “The lesson was difficult,” *Cronbach’s α* = 0.62), understanding (2 items; e.g., “I have a good understanding of the material,” *Cronbach’s α* = 0.78), effort, (2 items; e.g., “I put a lot of effort in the lesson”, *Cronbach’s α* = 0.91) interest (2 items; e.g., “I would like to see more of this type of lesson in classrooms,” *Cronbach’s α* = 0.74), enjoyment (2 items; e.g., “I enjoyed the lesson,” *Cronbach’s α* = 0.93), and motivation (2 items; e.g., “I felt motivated to understand the lesson,” *Cronbach’s α* = 0.77). Additionally, 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,” *Cronbach’s α* = 0.91), essential load (3 items; e.g., “I was trying to learn the main facts from the lesson,” *Cronbach's α* = 0.64), and generative load (3 items; e.g., "I was trying to make connections between the material and things I already know," *Cronbach's α* = 0.66). Eight items asked about four affective states, respectively: positive valence/high arousal (2 items; e.g., "I felt happy," *Cronbach's α* = 0.91), negative valence/high arousal (2 items; e.g., "I felt angry," *Cronbach's α* = 0.56), positive valence/low arousal (2 items; e.g., "I felt content," *Cronbach's α* = 0.59), and negative valence/low arousal (2 items; e.g., "I felt bored," *Cronbach's α* = 0.28).

The post-test (Appendix D) 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 4 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* = 0.77). Each of the four transfer questions were 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* = 0.65).

### ***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 x 18-foot lab room were used to track the participant's movements within the lesson. The desktop computer system consisted of a Dell desktop computer with a 24-inch monitor.

The brain-monitoring system was Advanced Brain Monitoring's B-Alert x10 mobile EEG system, which included 9 electrodes that corresponded to brain regions Fz, F3, F4, Cz, C3, C4, POz, P3, and P4 (https://www.advancedbrainmonitoring.com/xseries/x10). Two electrodes were also placed on the mastoid bone behind the ear as reference points. It also included 2 leads that attached to the collar bones to measure electrocardiography (ECG) data, which was used to calculate heart rate and HRV. The EEG and ECG electrodes were used with Synapse Conductive Electrode Cream, a non-abrasive NaCl-based cream.

To collect electrodermal activity (EDA), Biopac's wireless BioNomadix PPG/EDA module with the 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. This location on the palm has been validated as an accurate measure of EDA (van Dooren et al., 2012). The EDA electrodes were used with an electrode gel that was made of 0.5% saline added to a neutral base to create an isotonic, 0.05 molar NaCl, electrode paste.

The software used to acquire, process, and analyze 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 boredom (< .40), manageable or ideal workload (.40-.70), and high workload or cognitive overload ( >.70), which correlates with increased working memory load and difficulty level in mental arithmetic and other complex problem-solving tasks (Poythress et al., 2006). These classifications have also previously been used 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 4 classes, which included sleep onset, distraction, low engagement, and high engagement. Three benchmark tasks 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 three benchmark tasks included the short versions of the 3-choice Vigilance Task (3CVT), Visual Psychomotor Vigilance Task (VPVT), and Auditory Psychomotor Vigilance Task (APVT). Each task lasted approximately 3 minutes each. In the 3CVT, participants saw 1 of three shapes (triangle, upside down triangle, or square) sequentially. Their task was to press a "yes" key when they saw the triangle or a "no" key when they saw the upside-down triangle or square. Stimuli appeared frequently and required a high state of alertness during the first portion on the 3CVT, and the inter-stimulus intervals were extended to better identify individuals who were not able to remain engaged during the remaining portion. Performance on this task was used to map a *high engagement* cognitive state. The VPVT instructed the participants to press the Space bar every two seconds with their eyes open in time with the appearance of a large red circle displayed in the screen. This mapped onto the *low engagement* cognitive state. The APVT instructed the participants to press the Space bar every two seconds with their eyes closed. To help keep time, a tone was played every two seconds. This task mapped onto the *distraction* cognitive state. The *sleep onset* cognitive state was calculated using regression from the other three tasks.

EDA and ECG data were also processed using the AcqKnowledge software. 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 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 measure reflecting higher physiological arousal. Heart rate and heart rate variability (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.

Table 1 summarizes the major dependent measures collected in both experiments in this study: retention and transfer (as measures of learning outcomes), EEG workload and engagement (as measures of cognitive processing), and ECG and EDA (as measures of affective processing). It also summarizes a variety of self-report measures (e.g., perceived cognitive workload, enjoyment, interest, load, presence, and affect).

### ***Procedure***

Participants were run individually in a lab setting. They were randomly assigned to one of four biology lesson groups: IVRn, IVRp, PPTn, or PPTp. First, the participant was informed of the experimental procedure and signed a written consent form. 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 B-Alert EEG system was set up according to the manual, which included 9 electrodes held in place on the participant’s head with a Velcro headband, 2 electrode leads that were placed behind the participant’s ears on the mastoid bones as reference points, and 2 electrode leads which the participant placed on his or her collar bones to collect ECG data. Measurements were taken to check that the impedance of each EEG electrode was below 80 kΩ, ensuring a good connection between the sensor-scale surface. The Bionomadix transmitter was then placed on the participant’s wrist using a Velcro strap and the EDA electrodes were placed on the participant’s palm on his or her non-dominant hand near the thumb and pinky fingers. After the electrode set up, the participant completed the benchmark tasks that were used to calibrate his or her cognitive brain states.

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, 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 research would ask them a question out loud and wait for a response from the participant.

Before starting the lesson, a two-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. They were assisted in removing the electrodes and then were thanked for participating and debriefed on the purpose of the experiment. IRB approval was obtained and guidelines for research with human subjects were followed.

## Results

### *Did the Groups Differ on Basic Characteristics?*

A preliminary issue concerned whether the groups differed significantly on basic characteristics despite random assignment. A chi-square test 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 year in school, $F(3, 57) = 2.95$, $p = .040$. Post-hoc analyses showed that the PPTp group was significantly higher in 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 between groups.

Main effects of instructional media (i.e., IVR or PPT) and practice testing for 2 x2 ANOVAs and Analyses of covariance (ANCOVAs) were separated into the following sections according to the order of predictions. Estimated marginal means (EMMs) are reported for ANCOVAs. Effects sizes for main effects were reported as Cohen's *d*, which indicates a small effect size at $d = 0.2$, a medium effect size at $d = 0.5$, and a large effect size at $d = 0.8$ (Cohen, 1988). Effect sizes larger than 0.4 may be considered educationally relevant (Mayer, 2014). For comparisons between IVR and PPT conditions, a positive effect size favors IVR and a negative effect size favors PPT. For comparisons between practice and no practice testing, a positive effect size favors practice testing, and a negative effect size favors no practice testing. 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.

### *Did Instructional Media Affect Learning Outcomes?*

The primary issue in this study concerns whether instructional media affects performance on the retention and transfer tests. The distraction hypothesis predicts that learning in IVR should lead to lower retention and transfer scores than learning with PPT (prediction 1). The top portion of Table 2 shows the mean score (and standard deviation) for each of the four groups on the retention test and the transfer test for Experiment 1. A 2 (IVR vs. PPT) x 2 (practice testing vs. no practice testing) ANCOVA was run on retention and transfer test scores, with year in school as a covariate. On retention scores, the PPT groups ($EMM = 10.78$, $SE = 0.63$) scored higher than the IVR groups ($EMM = 8.99$, $SE = 0.64$), $F(1, 56) = 3.75$, $p = .058$, $d = -0.52$, but the difference did not reach statistical significance. The PPT groups ($EMM = 3.76$, $SE = 0.30$) scored significantly higher than the IVR groups on the transfer test ($EMM = 2.57$, $SE = 0.31$), $F(1, 56) = 7.11$, $p = .010$, $d = -0.71$. There were no significant interactions between 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.

### *Did Instructional Media Affect Cognitive Processes?*

Concerning cognitive processing during the lesson as measured by self-report ratings, the distraction hypothesis predicts that learning in IVR should lead to higher ratings of extraneous load and lower ratings of essential and generative load than learning with PPT (prediction 2a). The top section of Table 3 shows the mean rating (and standard deviation) for each of the four groups on extraneous, essential, and generative load. Two (IVR vs. PPT) x two (practice testing vs. no practice testing) ANOVAs were conducted on the extraneous, essential, and generative cognitive load subscales. As predicted, the IVR groups ($M$ = 3.52, $SD$ = 1.65) reported significantly higher extraneous 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 load ratings, $F$(1, 57) = 0.18, $p$ = .675, $d$ = 0.11, and the difference between the generative 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.

Concerning the cognitive processes during the lesson as measured by EEG, the distraction hypothesis also predicted that learning in IVR should lead to less ideal measures of cognitive workload and lower measures of cognitive engagement than learning with PPT (prediction 2b). The first section in Table 4 shows the mean rating (and standard deviation) for each of the four groups on the cognitive workload and engagement metrics for Experiment 1. Two (IVR vs. PPT) x 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 certain a 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-squares 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, $\chi 2$(1, N = 61) = 6.34, *p* = .012, *Cramer's V* = 0.32. Additionally, the IVR groups had a higher proportion of bored learners than PPT, $\chi 2$(1, N = 61) = 5.79, *p* = .012, *Cramer's V* = 0.31. 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 5 shows the frequencies of each type of cognitive workload learners in each condition experienced during the lesson. Overall, these findings support prediction 2b; there is evidence that learning in immersive virtual reality creates more distraction during learning than learning with a desktop.

### ***Did Instructional Media Affect Affective Processes?***

Concerning affective processing during learning based on self-report measures, the distraction hypothesis predicted that the IVR groups would report higher ratings of positive valance/high arousal emotions and lower ratings of negative valance/low arousal emotions as compared to the PPT groups (prediction 3a). The top portion of Table 6 shows the mean ratings (and standard deviations) of each group on four affective states during learning: positive valence/high arousal (happy or excited), positive valence/low arousal (calm or content), negative valence/high arousal (angry or frustrated), and negative valence/low arousal such as (sad or bored) in Experiment 1. ANOVAs were run on ratings of each of the 4 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, 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 hypothesis 3a and suggest that learning in virtual reality is emotionally arousing, particularly with positive emotions.

Concerning affective processing during learning, the distraction hypothesis predicted stronger indications of emotional arousal on biometric measures of heart activity and EDA for the IVR groups than the PPT groups (prediction 3b). The top portions of Tables 7 and 8 show the mean and standard deviation on three measures of electrodermal activity (EDA) and two measures of heart activity from electrocardiography (ECG). The three measures of EDA were skin conductance level from the tonic EDA signal (SCL), skin conductance responses per second (SCR), and peak-to-peak amplitude change from the phasic EDA signal; the two measures from ECG were heart rate and high-frequency heart rate variability (HRV). 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 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 peak-to-peak amplitude change compared to the PPT groups (*EMM* = 3.98, *SE* = .55), *F*(1, 55) = 2.37, *p* = .130, *d* = 0.41. However, there were several significant interactions between instructional media and practice testing on these measures. There was a significant interaction between instructional media and practice testing on tonic SCL, such that those in the IVR conditions with and without practice testing did not differ in tonic SCL, but those in the PPT conditions had increased tonic SCL with practice testing compared to no practice testing, *F*(1, 55) = 4.20, *p* = .045. There was a significant interaction between media and practice testing on SCR/sec, such that those in the IVR conditions with and without practice testing did not differ in SCR/sec, but those in the PPT conditions had increased SCR/sec with practice testing compared to no practice testing, *F*(1, 55) = 5.66, *p* = .021. The interaction between media and practice testing on peak-to-peak amplitude was not significant, *F*(1, 55) = 0.34, *p* = .561.

One extreme outlier was removed from the HRV analyses as it was greater than 4 standard deviations from the mean. The IVR groups (*EMM* = 6.23, *SE* = 7.70) did not differ significantly on HRV than the PPT groups (*EMM* = 13.37, *SE* = 0.38), *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 heart rate, *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 3b; learning in IVR did not create more affective processing as measured by physiological arousal than learning with conventional media.

### *Did Measures of Cognitive Processing Predict Learning Outcomes?*

The foregoing analyses provide some evidence that learning in immersive virtual reality leads to poorer learning outcomes, more distraction during learning, and more positive emotions during learning than learning with a desktop. In order to provide preliminary insights into the relations among cognitive processes and learning outcomes preceding mediation analyses, regression analyses were conducted. First, a multiple regression analysis was conducted with extraneous load, essential load, and generative load as predictors for retention scores, as shown in the top portion of Table 9. Together the three predictors explained approximately 18% of retention scores, $R^2 = .18$, $F(3, 57) = 4.12$. $p = .010$. Self-reported extraneous load explained a significant amount of unique variance in retention score, $\beta = -.39$, $p = .003$, $sr^2 = .14$, indicating that retention scores were higher for those who reported lower extraneous cognitive load. The three cognitive load predictors also explained approximately 15% of the variance in transfer scores, $R^2 = .15$, $F(3, 57) = 3.37$. $p = .024$, as shown in the top section of Table 10. Self-reported extraneous load explained a significant amount of unique variance in transfer score, $\beta = -.30$, $p = .020$, $sr^2 = .09$, indicating that transfer scores were higher for those who reported lower extraneous cognitive load.

Predictors of learning outcomes from the EEG based cognitive metrics were also examined. A multiple regression analysis was conducted with probabilities for high engagement, low engagement, distraction, and sleep onset, and workload score as predictors for retention and transfer scores. As seen in the second portion of Table 9, the five predictors did not significantly predict retention scores, $R^2 = .09$, $F(5, 55) = 1.04$, $p = .407$. However, the second section of Table 10 shows that the cognitive activity predictors significantly explained approximately 21% of the variance in transfer scores, $R^2 = .21$, $F(5, 55) = 2.93$, $p = .020$. Probabilities of low engagement, $\beta = -.85$, $p = .004$, $sr^2 = .13$, distraction, $\beta = -.48$, $p = .012$, $sr^2 = .10$, and sleep onset, $\beta = -.55$, $p = .004$, $sr^2 = .13$, explained a significant amount of unique variance in transfer score, indicating that transfer scores were higher for those who had a lower probability of being in a low engagement, distraction, or sleep onset cognitive state.

### *Did Measures of Affective Processing Predict Learning Outcomes?*

Similarly, in order to provide preliminary insights into the relations among affective processes and learning outcomes preceding mediation analyses, regression analyses were conducted. To examine the predictors of learning outcomes from self-reported affective states, a multiple regression analysis was conducted with the 4 categories of affective states as predictors for retention and transfer scores, as shown in the bottom two sections of Tables 9 and 10. Affective states together did not significantly predict retention scores, $R^2 = .04$, $F(4,56) = 0.52$, $p = .720$, or transfer scores, $R^2 = .09$, $F(4,56) = 1.31$, $p = .277$. Predictors of learning outcomes from the EDA and ECG metrics were also examined. EDA and ECG measures did not significantly predict retention scores, $R^2 = .12$, $F(5, 54) = 1.47$, $p = .214$, or transfer scores, $R^2 = .08$, $F(5, 54) = 0.93$, $p = .469$.

### ***Did Cognitive or Affective Processing Mediate the Relationship between Instructional Media and Learning Outcomes?***

The ANOVAs and regression analyses suggest that cognitive and affective processes are related to both instructional media and learning outcomes. To explore whether cognitive and/or affective processing during the lesson explain the relationship between instructional media and learning outcomes, a parallel mediation was conducted, as shown in Figure 1a. To reduce the number of mediators in the model, extraneous cognitive load, high engagement, positive/high arousal emotions, and heart rate variability were selected as they best represented each category of self-report and biometric measurements of cognitive and affective processing during the lesson. Figure 1b shows this updated model. It was predicted that these measures would mediate the relationship between instructional media and learning outcomes; as immersion increased (comparing desktop to IVR), measures of emotional/physiological arousal and cognitive distraction would increase, which would decrease performance on retention and transfer tests (prediction 4). The mediation analysis was conducted with the PROCESS macro in SPSS (Hayes, 2018) based on 5000 bootstrapped samples. Instructional media was coded as a categorical variable with the IVR lesson as 0 and the desktop lesson as 1.

Figure 2 shows the two mediation models run on retention and transfer scores in Experiment 1. The total effect of instructional media on retention scores was significant [*b* = -1.93, *SE* = 0.83, *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* = 0.02, 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* = 0.67, 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* = 0.22, 95% CI (-0.69, -0.17)] and heart rate variability [*b* = 0.16, *SE* = 0.15, 95% CI (-0.03, 0.53)] were not significant.

For transfer scores, the total effect of instructional media was significant [*b* = -1.45, *SE* = 0.40, *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* = 0.57, *p* = .010]. The indirect paths of instructional media on transfer scores through extraneous cognitive load [*b* = -0.16, *SE* = 0.16, 95% CI (-0.52, 0.13)], high engagement [*b* = -0.07, *SE* = 0.10, 95% CI (-0.28, 0.14)], positive, high arousal emotions [*b* = 0.32, *SE* = 0.36, 95% CI (-0.36, 1.07)], and heart rate variability [*b* = -0.01, *SE* = 0.06, 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, which partially supports prediction 4.

It may be noted that the direct effects in both models were still significant after accounting for the mediators, although reduced from the total effect. The casual steps approach (Baron & Kenny, 1986) would argue that this indicates only *partial* mediation (as opposed to *full* mediation). However, more modern approaches suggest that the focus should be moved toward examining indirect effects, rather than the total or direct effects, as they provide more insight on understanding the mechanisms by which the independent variable's effect operates on the dependent variable. Therefore, determining which indirect effects differ from 0 are the focus of these analyses, rather than a comparison of the total and direct effects (Hayes, 2018).

### ***Did Practice Testing Affect Learning Outcomes?***

Contrary to the predictions of the generative learning strategy hypothesis (prediction 5), 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.

### ***Did Practice Testing Affect Cognitive Processes?***

In partial support of the predictions of the generative learning hypothesis (prediction 6a), those who engaged in practice testing (*M* = 5.49, *SD* = 0.93) rated their essential load as significantly higher than those who did not engage in practice testing (*M* = 4.94, *SD* = 1.14), *F*(1, 57) = 4.36, *p* = .041, *d* = 0.55. However, there were no differences in ratings of extraneous processing between practice testing (*M* = 3.08, *SD* = 1.53) and no practice testing (*M* = 2.83, *SD* = 1.55), *F*(1, 57) = 0.39, *p* = .533, *d* = 0.17, and the practice testing groups

(*M* = 5.06, *SD* = 1.09) and no practice testing groups (*M* = 5.31, *SD* = 1.15) also did not differ on generative processing, *F*(1, 57) = 0.78, *p* = .381, *d* = -0.23.

Concerning the cognitive process as measured by EEG, the generative learning strategy hypothesis also predicted that students in the practice testing condition should exhibit levels of lower workload and higher levels of engagement during learning than students in the no practice testing condition (prediction 6b). However, this prediction was not supported as there were no significant differences between the practice (*EMM* = .41, *SE* = .02) and no practice groups (*EMM* = .39, *SE* = .02) for high engagement, *F*(1, 56) = 0.28, *p* = .601, *d* = 0.14, and no differences between the practice (*EMM* = .42, *SE* = .02) and no practice groups (*EMM* = .42, *SE* = .02) for low engagement, *F*(1, 56) = 0.01, *p* = .923, *d* < 0.001. For distraction, the difference between practice (*EMM* = .05, *SE* = .02) and no practice groups (*EMM* = .08, *SE* = .02) was not significant, *F*(1, 56) = 1.28, *p* = .263, *d* = 0.30, and the practice groups (*EMM* = .10, *SE* = .02) and no practice groups (*EMM* = .10, *SE* = .02) also did not differ on sleep onset, *F*(1, 56) = 0.04, *p* = .838, *d* = 0.06. Additionally, the proportion of learners in cognitive workload states of boredom, $\chi 2$(1, N = 61) = 2.79, *p* = .095, *Cramer's V* = 0.21, ideal workload, $\chi 2$(1, N = 61) = 0.51, *p* = .470, *Cramer's V* = 0.09, and overload, $\chi 2$(1, N = 61) = 2.07, *p* = .150, *Cramer's V* = 0.18, did not differ.

### ***Did Practice Testing Affect Affective Processes?***

Concerning the effect of practice testing on affective processing, it was predicted that practice testing would not differ between practice testing and no practice testing groups (prediction 7a). Those in the practice testing groups (*M* = 4.03, *SD* = 1.78) did not significantly differ from no practice testing groups (*M* = 4.65, *SD* = 1.52) in their self-report ratings of positive/high arousal emotions, *F*(1, 57) = 3.57, *p* = .064, *d* = 0.50. Practice (*M* = 4.85, *SD* = 1.18) and no practice testing (*M* = 5.02, *SD* = 1.07) groups also did not differ in their ratings of positive/low arousal emotions, *F*(1, 57) = 0.31, *p* = .580, *d* = 0.14. For ratings of negative/high arousal emotions, the practice testing (*M* = 1.40, *SD* = 0.71) and no practice testing groups (*M* = 1.15, *SD* = 0.37) did not differ, *F*(1, 57) = 3.19, *p* = .079, *d* = 0.47, and the practice testing (*M* = 1.92, *SD* = 1.08) and no practice testing groups (*M* = 1.87, *SD* = 0.91) did not differ on negative/low arousal emotions, *F*(1, 57) = 0.05, *p* = .815, *d* = 0.06. These findings support prediction 7a.

Similarly, prediction 7b was that groups would also not differ in physiological measures of emotional arousal between practice and no practice testing groups. The practice testing groups (*EMM* = 11.48, *SE* = .41) had significantly higher tonic SCL than the no practice testing groups (*EMM* = 10.13, *SE* = .40), *F*(1, 55) = 5.49, *p* = .023, *d* = 0.63, and the practice testing groups (*EMM* = 5.94, *SE* = .54) had significantly higher peak-to-peak amplitude than the no practice testing groups (*EMM* = 3.27, *SE* = .54), *F*(1, 55) = 12.10, *p* = .001, *d* = 0.94. However, practice testing groups (*EMM* = .09, *SE* = .004) did not significantly have more SCR/sec than no practice testing groups (*EMM* = .07, *SE* = .004), *F*(1, 55) = 3.49, *p* = .067, *d* = 0.51.

Additionally, practice testing groups (*EMM* = 84.74, *SE* = .85) had higher heart rate during the lesson than no practice test groups (*EMM* = 82.35, *SE* = .84), but this did not reach significance, *F*(1, 55) = 3.95, *p* = .052, *d* = 0.54. Practice testing groups (*EMM* = 9.19, *SE* = 7.24), did not differ in HRV than no practice testing groups (*EMM* = 10.41, *SE* = 7.24), *F*(1, 54) = 0.01, *p* = .907, *d* < 0.001. Although self-report measures suggest adding practice testing to lessons did not affect learners' emotional arousal, these findings show that do increase physiological arousal, which do not support predict 7b. Rather, these findings are more indicative of the alternative hypothesis that oral practice testing questions are physiologically arousing, perhaps because they are stressful. Thus, there is some evidence that being asked to answer questions orally to an experimenter at various points in the lesson created greater physiological arousal than not being asked to answer questions, which could have caused some distraction for the learner.

### *Did Instructional Media or Practice Testing Affect Presence or Other Self-Report Measures?*

Multiple 2 (IVR vs. PPT) x 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 ($M = 3.98$, $SD = 1.64$), $F(1, 57) = 20.59$, $p < .001$, $d = 1.20$, and the IVR groups ($M = 2.88$, $SD = 1.08$) also reported significantly lower perceived difficulty than the PPT groups ($M = 3.61$, $SD = 1.40$), $F(1, 57) = 5.54$, $p =.022$, $d = 0.63$. No other comparisons between media significantly differed in perceived 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. Average ratings and standard deviations for each measure are shown in the Table 13. As expected, students felt more presence, enjoyed the experience more, and perceived the lesson as easier in immersive virtual reality than with conventional media; however, liking the experience did not translate into better learning processes or outcomes.

## Discussion

The results from Experiment 1 showed that students performed better on transfer tests when they learned with conventional media in the form of a desktop lesson rather than in IVR. Second, concerning cognitive processes, students who learned with the desktop lesson reported lower levels of extraneous cognitive load based on self-report measures and exhibited higher levels of engagement based on EEG measures than students who learned in IVR. Third, concerning affective processes, students who learned with the desktop lesson reported lower levels of positive/high arousal emotion during learning and higher levels of negative/low arousal emotions during learning based on self-report measures than students who learned in IVR. The relationship between instructional media was mediated by some measures of cognitive and affective processes, particularly self-reported cognitive load and positive/high arousal emotions, suggesting that instructional media that causes cognitive or affective distractions leads to poorer learning outcomes. Overall, the results provide some evidence supporting the distraction hypothesis, which was derived from the Cognitive Theory of Multimedia Learning. Learning academic content in immersive virtual reality can create emotional arousal and distraction that leads to poorer transfer test performance.

The results of this study also show that adding practice testing does not significantly improve post-test performance or cognitive processes, but does tend to increase physiological arousal. The increase in physiological arousal may be due to the stress of having to verbally answer questions in front of the experimenter. Finally, although students do not learn as well in immersive virtual reality, they enjoy it more, feel a greater sense of presence, and perceive it as an easier media in which to learn as compared to learning the same material from a PPT. This pattern confirms the adage that liking is not learning (Sung & Mayer, 2012).

# Chapter III: Experiment 2

Similar to Experiment 1, the goal of Experiment 2 was to compare learning outcomes of a history lesson between IVR and a video lesson on a desktop. Additionally, the relationships between instructional media, cognitive processes, affective processes, and learning outcomes were explored.

## Method

### *Participants and Design*

The participants were 80 undergraduate students aged 17 through 23 years ($M$ = 19.28, $SD$ = 1.25) recruited through the Psychology Paid Subject Pool through the Sona website at the University of California, Santa Barbara. There were 19 males and 61 females. Participants were compensated 20 dollars for completion of the experiment. Participants were randomly assigned to one of four groups in a 2 (Instructional media: IVR, Video) x 2 (practice testing vs. no practice testing) between-subjects design, and there were 20 participants in each group.

### *Materials*

The instructional materials consisted of four versions of an academic history lesson: IVR, IVR with practice questions (IVRp), Video, and Video with practice questions (Videop). The academic lesson used was taken from *Kokoda VR,* as shown in Figure 6 (Australian Broadcasting Company, 2018). The lesson explained the Kokoda Track campaign, which was a series of battles during the Pacific War of World War II between the Japanese and Australians over the Australian territory, Papua (present-day Papua-New Guinea). It included real historical interviews and videos, as well as 3D simulations of locations and artifacts from the battles and simulations of dialogue and war scenes among military personnel and other parties involved. In the IVR lesson, the learner could move around the lesson and pick up and interact with some objects throughout the lesson. The Video version of the lesson was a 360-degree YouTube video playable on a desktop computer, in which the learner could move his or her perspective angle of view, but could not interact with objects in the lesson. Chapters 1, 2, 3, 9, 10, and 11 were used for the lesson, and both the IVR and Video versions lasted approximately 22 minutes. In the IVRp and Videop conditions, the lessons were segmented into 6 sections consisting of 1 chapter each; at the end of each chapter, the researcher verbally asked the participant a question pertaining to the previously viewed section (Appendix G). Participants were given unlimited time to answer, and no feedback was given after each answer.

The paper-based materials were the same as Experiment 1, except that the prior knowledge survey and post-test focused on history, rather than biology (Appendix B). The prior history knowledge survey asked about experience in previous history classes, activities (e.g., "I enjoy watching non-fictional history documentaries"), and a self-reported rating of general WWII knowledge. A prior knowledge score was created by adding the number of history-related classes the learner took in high school and college, the number of items checked on the activity checklist, and a score between 0 and 3 on the learner's rating of knowledge of WWII. Similar to Experiment 1, a pre-test on the content of the history lesson was not used to prevent a testing effect.

The post-lesson questionnaire was the same as the one used in Experiment 1 with the following reliabilities for each subscale: presence (*Cronbach's α* = 0.87), difficulty (*Cronbach's α* = 0.75), understanding (*Cronbach's α* = 0.81), effort (*Cronbach's α* = 0.91), interest (*Cronbach's α* = 0.81), enjoyment (*Cronbach's α* = 0.94), motivation (*Cronbach's α* = 0.58), extraneous load (*Cronbach's α* = 0.87), essential load (*Cronbach's α* = 0.69), and generative load (*Cronbach's α* = 0.29), positive valence/high arousal (*Cronbach's α* = 0.82), negative valence/high arousal (*Cronbach's* $\alpha = 0.70$), positive valence/low arousal (*Cronbach's* $\alpha = 0.60$), and negative valence/low arousal (*Cronbach's* $\alpha = 0.29$).

Similar to Experiment 1, the post-test consisted of 16 retention questions in the form of multiple-choice items (e.g., “What were the Australian New Guinea Administrative Unit responsible for? (a) Organizing labor and supplies, (b) Navigating the rough terrain, (c), Housing the Australian troops, (d) All of the above”) and 4 open-ended transfer questions (e.g., “Why was Kokoda an important place to capture?”; Appendix E). A retention score was calculated from the number of correct answers with a total of 16 points (*Cronbach's* $\alpha = 0.54$). Transfer questions had between 3 and 6 correct answers, for a total of 17 points (*Cronbach's* $\alpha = 0.76$).

### *Apparatus*

The same virtual reality system (HTV Vive), desktop computer system (Dell computer), brain-monitoring system (B-Alert x10), and biometric-monitoring system (BioNomadix) from Experiment 1 were used for Experiment 2.

### *Procedure*

Participants were randomly assigned to one of the four conditions: IVR, IVRp, Video, or Videop. The same procedure was used as in Experiment 1, except for the desktop videos. Participants in the Video and Videop conditions were given instructions on how to view the lesson (e.g., how to change their angle of view using a click-and-drag movement). Those in the Videop condition were given additional instructions to pause at the end of each of the six video lessons, during which the researcher would ask him or her a verbal question, wait for a response from the participant, and then continue onto the next chapter in the video lesson.

## Results

### *Did the Groups Differ on Basic Characteristics?*

A preliminary issue concerned whether the groups were equivalent on basic characteristics before the start of the study. In order to determine whether the 4 groups had pre-existing differences, ANOVAs were run on age, year in school, and background knowledge, and a chi-square analysis was run on proportion of men and women. The groups did not differ significantly in age, $F(3, 76) = 0.95$, $p = .423$, year in school, $F(3, 76) = 0.54$, $p = .655$, background knowledge, $F(3, 76) = 0.30$, $p = .822$, or gender, $\chi^2(3, N = 80) = 2.97$, $p = .397$.

Similar to the results in Experiment 1, the main effects of instructional media (IVR or Video) and practice testing (practice testing or no practice testing) for 2x2 ANOVAs and ANCOVAs are reported in the following sections in the order of the predictions. Estimated marginal means are reported for ANCOVAs. Effects sizes for main effects were reported as Cohen's *d*, which indicates a small effect size at $d = 0.2$, a medium effect size at $d = 0.5$, and a large effect size at $d = 0.8$ (Cohen, 1988). Effect sizes larger than 0.4 may be considered educationally relevant (Mayer, 2014). For comparisons between IVR and Video conditions, a positive effect size favors IVR and a negative effect size favors Video. For comparisons between practice and no practice testing, a positive effect size favors practice testing, and a negative effect size favors no practice testing.

### *Did Instructional Media Affect Learning Outcomes?*

Prediction 1 was that the Video groups would have higher performance on retention and transfer tests than the IVR groups. The bottom portion Table 2 shows the means (and standard deviations) of the 4 groups on retention score and transfer score for Experiment 2. A 2 (IVR vs. Video) x 2 (practice testing vs. no practice testing) ANOVA was run on retention and transfer scores. On retention scores, there was no significant difference between students who learned with IVR (*M* = 9.15, *SD* = 2.73) and those who learned with Video (*M* = 9.53, *SD* = 2.54), *F*(1, 76) = 0.42, *p* = .521, *d* = -0.14. On transfer scores, students who viewed the Video lesson (*M* = 4.09, *SD* = 2.34) scored significantly higher than those who viewed the IVR lesson (*M* = 3.06, *SD* = 2.05), *F*(1, 76) = 4.25, *p* = .043, *d* = 0.47. There were no significant interactions between instructional media and practice testing for retention score, *F*(1, 76) = 1.56, *p* = .216, or transfer score, *F*(1, 76) = 0.21, *p* = .652. These results partially support the first prediction that learners learn more from a desktop lesson than an IVR lesson, but only for transfer tests.

### ***Did Instructional Media Affect Cognitive Processes?***

For cognitive processes, the distraction hypothesis predicted that learners would report higher extraneous cognitive load and lower essential and generative load after the IVR lesson compared to the Video lesson (prediction 2a). The second section of Table 3 shows the mean ratings (and standard deviations) for each group on extraneous, essential, and generative load. Two (IVR vs. Video) x two (practice testing vs. no practice testing) ANOVAs revealed that ratings of extraneous cognitive load did not differ significantly between the IVR (*M* = 3.78, *SD* = 1.67) and video groups (*M* = 3.20, *SD* = 1.32) for extraneous load, *F*(1, 76) = 3.12, *p* = .081, *d* = 0.40. Additionally, the IVR groups (*M* = 5.36, *SD* = 1.23) did not have different ratings on essential load than the Video groups (*M* = 5.47, *SD* = 0.93), *F*(1, 76) = 0.20, *p* = .656, *d* = -0.11, and the difference between the IVR (*M* = 5.07, *SD* = 1.07) and Video groups (*M* = 4.83, *SD* = 0.95) on generative load was not significant, *F*(1, 76) = 1.12, *p* = .294, *d* = 0.24. There were no significant interactions between instructional media and practice testing on extraneous load ratings, *F*(1, 76) = 0.50, *p* = .482, essential load ratings, *F*(1, 76) = 1.45, *p* = .232, or generative load ratings, *F*(1, 76)

= 0.00 *p* = .971. These findings do not support the prediction 2a; IVR did not cause increased extraneous cognitive load.

Prediction 2b was that the desktop lesson would result in higher probabilities of learners being in a high cognitive engagement state and a higher proportion of learners with ideal cognitive workload. The second section of Table 4 shows the mean probabilities (and standard deviations) of each cognitive engagement state for all conditions during the lesson for Experiment 2. Two by two ANCOVAs were conducted on cognitive engagement scores, with baseline measures as covariates. As predicted, students in the Video conditions (*EMM* = 0.42, *SE* = 0.03) exhibited a higher probability of being in a high engagement cognitive state than students in the IVR conditions (*EMM* = 0.35, *SE* = .03), *F*(1, 75) = 4.14, *p* = .045, *d* = -0.47. There were no differences between the IVR (*EMM* = 0.38, *SE* = .02) and Video groups (*EMM* = 0.41, *SE* = .02) on low engagement, *F*(1, 75) = 0.64, *p* = .427, *d* = -0.18. The IVR groups (*EMM* = 0.12, *SE* = .02) also did not differ on distraction probability compared to the Video groups (*EMM* = 0.10, *SE* = .02), *F*(1, 75) = 0.51, *p* = .476, *d* = 0.17. Finally, for sleep onset, there was no difference between the IVR (*EMM* = 0.10, *SE* = .01) and Video groups (*EMM* = 0.10, *SE* = .01), *F*(1, 75) = 1.42, *p* = .237, *d* = 0.29. There were no significant interactions between instructional media and practice testing on high engagement, *F*(1, 75) = 0.52, *p* = .475, low engagement, *F*(1, 75) = 0.07, *p* = .789, distraction, *F*(1, 75) = 0.13, *p* = .721, or sleep onset, *F*(1, 75) = 0.16, *p* = .695.

Chi-square analyses were run on the proportion of learners classified in the three states of cognitive workload (boredom, ideal, and overload) between IVR and Video groups; frequencies for each group are shown in Table 5. The proportion of participants with ideal workload during the lesson did not differ between the IVR groups and Video groups, $\chi_2$(31, N = 80) = 0.45, *p* = .501, *Cramer's V* = 0.08. Additionally, the proportions of bored, $\chi_2$(31, N = 80) = 1.27, *p* = .260, *Cramer's V* = 0.13, and overloaded participants, $\chi2$(31, N = 80) = 2.05, *p* = .152, *Cramer's V* = 0.16, did not differ between IVR and Video groups. These findings partially support prediction 2b. There is some EEG-based evidence that learning with a desktop video lesson is less distracting than learning with IVR as participants were in a higher cognitive engagement state.

### ***Did Instructional Media Affect Affective Processes?***

Prediction 3a stated that learners would experience more positive/high arousal emotions and less negative/low arousal emotions in the IVR lesson compared to the video lesson. The bottom section of Table 6 shows the mean ratings (and standard deviations) of each group on four affective states during learning in Experiment 2. As predicted, 2 x 2 ANOVAs revealed that the IVR groups (*M* = 4.73, *SD* = 1.43) reported significantly higher ratings of positive valence/high arousal emotions than the Video groups (*M* = 3.58, *SD* = 1.37), *F*(1, 76) = 13.20, *p* = .001, *d* = 0.83, and the IVR groups (*M* = 2.35, *SD* = 1.14) reported significantly lower ratings of negative valence/low arousal emotions than the Video groups (*M* = 3.14, *SD* = 1.07), *F*(1, 76) = 9.99, *p* = .002, *d* = -0.72. Similarly, the IVR groups (*M* = 1.89, *SD* = 1.18) also reported lower ratings than the Video groups (*M* = 2.80, *SD* = 1.52) on negative valence/high arousal emotions, *F*(1, 76) = 8.78, *p* = .004, *d* = -0.68. IVR groups (*M* = 4.50, *SD* = 1.26) did not differ significantly in ratings of positive valence/low arousal emotions than the video groups (*M* = 4.64, *SD* = 1.09), *F*(1, 76) = 0.27, *p* = .607, *d* = -0.13. There were no significant interactions between instructional media and practice testing for ratings of positive/high arousal emotions, *F*(1, 76) = 0.23, *p* = .637, negative/high arousal emotions, *F*(1, 76) = 0.28, *p* = .599, positive/low arousal emotions, *F*(1, 76) = 0.02, *p* = .888, or negative/low arousal emotions, *F*(1, 76) = 0.27, *p* = .607. This pattern of results supports prediction 3a of the distraction hypothesis as IVR caused higher positive emotional arousal than the video.

The distraction hypothesis predicted that learners would have higher levels of physiological arousal in the IVR lesson compared to the video lesson (prediction 3b). Tables 7 and 8 show the mean and standard deviation on three measures of electrodermal activity (EDA) and two measures of heart activity from electrocardiography (ECG), respectively. Corresponding 2 x 2 ANCOVAs were run, with movement during the lesson and respective baseline measures as covariates. The IVR groups (*EMM* = 9.77, *SE* = .71) did not differ significantly in average SCL from the Video groups (*EMM* = 11.20, *SE* = .71), *F*(1, 74) = 1.64, *p* = .204, *d* = -0.30. SCRs per second also did not differ between the IVR conditions (*EMM* = .08, *SE* = .01) and Video conditions (*EMM* = .07, *SE* = .01), *F*(1, 74) = 0.05, *p* = .819, *d* = -0.06. For peak-to-peak amplitude, IVR (*EMM* = 6.94, *SE* = 1.31) did not cause a greater change than the Video (*EMM* = 7.56, *SE* = 1.31), *F*(1, 74) = 0.09, *p* = .766, *d* = -0.06. The IVR groups had significantly higher heart rate (*EMM* = 89.06, *SE* = 1.04) during the lesson than the Video groups (*EMM* = 83.23, *SE* = 1.04), *F*(1, 74) = 12.51, *p* = .001, *d* = 0.82. However, the IVR groups (*EMM* = 7.08, *SE* = 2.10) did not differ from the Video groups (*EMM* = 8.88, *SE* = 2.10) in HRV, *F*(1, 74) = 0.29, *p* = .591, *d* = -0.13. There was a significant interaction between instructional media and practice testing on peak-to-peak amplitude, *F*(1, 74) = 4.73, *p* = .033, such that within the Video group, practice testing largely increased the peak-to-peak amplitude change, whereas within the IVR group, practice testing did not. There were no significant interactions between instructional media and practice testing on SCL, *F*(1, 74) = 1.61, *p* = .209, SCR, *F*(1, 74) = 0.39, *p* = .535, heart rate, *F*(1, 74) = 0.10, *p* = .754, or HRV, *F*(1, 74) = 2.92, *p* = .0.92. These findings are partially consistent with prediction 3b, showing that IVR caused more physiological arousal, particularly with heart rate measures.

### *Did Cognitive or Affective Processing Predict Learning Outcomes?*

Multiple regression analyses, similar to Experiment 1, were conducted to examine whether cognitive and affective processes predicted learning outcomes prior to running mediation analyses. There were no significant cognitive or affective predictors for retention scores, as shown in Table 10; self-reported cognitive load, $R^2 = .04$, $F(3, 76) = 0.91$, $p = .439$, EEG-based cognitive engagement and workload, $R^2 = .04$, $F(5, 74) = 0.57$, $p = .726$, self-reported emotional states, $R^2 = .07$, $F(4, 75) = 1.43$, $p = .234$, and physiological measures, $R^2 = .04$, $F(4, 74) = 0.54$, $p = .743$, did not predict performance on the retention test.

For transfer scores, the three cognitive load predictors, $R^2 = .08$, $F(3, 76) = 2.05$, $p = .114$, five measures of EEG-based engagement and workload scores, $R^2 = .03$, $F(3, 74) = 0.52$, $p = .762$, and five physiological measures of emotions, $R^2 = .07$, $F(3, 74) = 1.18$, $p = .326$, did not explain a significant amount of the variance in scores as shown in Table 11. Self-reported emotions significantly explained 12% of the variance in transfer scores, $R^2 = .12$, $F(3, 75) = 2.59$, $p = .043$. Self-reported positive, low arousal emotions, $\beta = .26$, $p = .025$, $sr^2 = .16$, and negative, low arousal emotions, $\beta = .30$ $p = .032$, $sr^2 = .25$, explained a significant amount of unique variance in transfer score, indicating that transfer scores were higher for those who reported high low arousal emotions or that low arousal emotions are less distracting to learning processes.

### *Did Cognitive or Affective Processing Mediate the Relationship between Instructional Media and Learning Outcomes?*

Similar to Experiment 1, to test the prediction that cognitive and affective processing mediated the relationship between instructional media and learning (prediction 4), a parallel mediation analysis explored whether extraneous load, EEG-based high engagement scores, positive/high arousal emotions, and HRV mediated the relationship. The mediation was conducted with the PROCESS macro in SPSS (Hayes, 2018) based on 5000 bootstrapped samples. Instructional media was coded as a categorical variable with the IVR lesson as 0 and the desktop lesson as 1. Figure 3 shows the unstandardized regression coefficients for the pathways in the models for retention and transfer. Neither the total effect of instructional media on retention scores [*b* = -.38, *SE* = 0.59, *p* = .527], nor the direct effect of instructional media on retention scores after including the mediators were significant [*b* = -.78, *SE* = .64, *p* = .238]. However, one significant indirect effect emerged; self-reported positive/high arousal emotions [*b* = 0.48, *SE* = 0.26, 95% CI (0.02, 1.06)] significantly mediated the effect of instructional media on retention scores. Those in the IVR conditions reported more positive, high arousal emotions (e.g., happy, excited), which caused lower overall retention scores than the Video conditions. Self-reported extraneous load [*b* = -0.04, *SE* = 0.15, 95% CI (-0.41, 0.23)], the high engagement EEG metric [*b* = 0.02, *SE* = 0.12, 95% CI (-0.18, 0.32)] and heart rate variability [*b* = -0.07, *SE* = 0.10, 95% CI (-0.32, 0.04)] did not significantly mediate the relationship between instructional media and retention scores.

For transfer scores, the total effect of instructional media was significant [*b* = -1.03, *SE* = 0.49, *p* = .041], and the direct effect of instructional media on transfer scores after including the mediators was no longer significant [*b* = -0.95, *SE* = 0.32, *p* = .083], suggesting a mediation of the overall model. However, no individual mediator significantly mediated the relationship. The indirect effects of extraneous cognitive load [*b* = -0.13, *SE* =

0.13, 95% CI (-0.44, 0.08)], high engagement [*b* = 0.08, *SE* = 0.10, 95% CI (-0.16, 0.26)], positive, high arousal emotions [*b* = 0.13, *SE* = 0.22, 95% CI (-0.21, 0.68)], and heart rate variability [*b* = -.09, *SE* = 0.08, 95% CI (-0.28, 0.05)] did not significantly mediate the relationship between instructional media and transfer scores. These results suggest that affective processes mediate the relationship between instructional media and learning outcomes, particularly for retention tests, which partially supports prediction 4.

### ***Did Practice Testing Affect Learning Outcomes?***

Scores on the retention test did not differ significantly for students who engaged in practice testing (*M* = 9.83, *SD* = 2.40) versus students who did not (*M* = 8.85, *SD* = 2.79), *F*(1, 76) = 2.81, *p* = .098, *d* = 0.39; and scores on the transfer test did not differ significantly for students who engaged in practice testing (*M* = 3.68, *SD* = 2.27) versus students who did not (*M* = 3.48, *SD* = 2.25), *F*(1, 76) = 0.16, *p* = .689, *d* = 0.09. These findings do not support the main prediction of the generative learning hypothesis (prediction 5) that practice testing would cause improvements in learning outcomes.

### ***Did Practice Testing Affect Cognitive Processes?***

It was hypothesized that adding practice testing to the lesson would result in higher ratings of essential and generative load and lower ratings of extraneous load than no practice testing (prediction 6a). Contrary to this prediction, students in the practice testing groups reported significantly higher extraneous cognitive load (*M* = 3.84, *SD* = 1.30) than those in the no practice testing groups (*M* = 3.15, *SD* = 1.67), *F*(1, 76) = 4.49, *p* = .037, *d* = 0.49. The practice testing (*M* = 5.56, *SD* = 0.89) and no practice testing groups (*M* = 5.27, *SD* = 1.24) did not differ in their ratings of essential load, *F*(1, 76) = 1.45, *p* = .232, *d* = 0.28, and ratings of generative load did not differ between practice (*M* = 5.03 *SD* = 1.01) and no practice groups (*M* = 4.86, *SD* = 1.03), *F*(1, 76) = 0.58, *p* = .447, *d* = 0.18. These results do not support the generative learning hypothesis (prediction 6a), but do support the alternative hypothesis that this form of practice testing may have been distracting.

Concerning the cognitive processes as measured by EEG, the generative learning strategy hypothesis also predicted that students in the practice testing conditions should exhibit more learners with ideal levels of workload and higher levels of engagement during learning than students in the no practice testing conditions (prediction 6b). However, this hypothesis was not supported as there were no significant differences between the practice (*EMM* = .39, *SE* = .03) and no practice groups (*EMM* = .38, *SE* = .03) in their probabilities of being in a high engagement cognitive state, $F(1, 75) = 0.18$, $p = .675$, $d = 0.09$. Additionally, the practice groups (*EMM* = .39, *SE* = .02) and no practice groups (*EMM* = .40, *SE* = .02) did not differ in their probabilities of being in a low engagement cognitive state, $F(1, 75) = 0.15$, $p = .703$, $d = -0.09$. For distraction probabilities, the practice (*EMM* = .11, *SE* = .02) and no practice groups (*EMM* = .11, *SE* = .02) did not differ, $F(1, 75) = 0.01$, $p = .936$, $d < 0.001$, and probabilities for sleep onset between the practice (*EMM* = .08, *SE* = .01) and no practice groups (*EMM* = .09, *SE* = .01) also did not differ, $F(1, 75) = 0.74$, $p = .393$, $d = -0.20$. Additionally, the proportion of learners in cognitive workload states of boredom, $\chi 2(1, N = 80) = 2.49$, $p = .115$, *Cramer's V* = 0.18, ideal workload, $\chi 2(1, N = 80) = 1.26$, $p = .262$, *Cramer's V* = 0.13, and overload, $\chi 2(1, N = 80) = 2.05$, $p = .152$, *Cramer's V* = 0.16, did not differ.

### ***Did Practice Testing Affect Affective Processes?***

It was predicted that practice testing would not change emotional states during the lesson compared to not including practice testing questions (prediction 7a). Concerning the effect of practice testing on affective processing, students in the practice testing groups did not significantly differ from the no practice testing groups in their self-report ratings of emotions during the lesson. Practice (*M* = 4.18, *SD* = 1.43), and no practice testing groups (*M* = 4.13, *SD* = 1.60) did not differ in ratings of positive/high arousal emotions, *F*(1, 76) = 0.03, *p* = .875, *d* = 0.00. For positive/low arousal emotions, practice (*M* = 4.64, *SD* =1.24) and no practice groups (*M* = 4.50, *SD* = 1.12) also did not differ, *F*(1, 76) = 0.27, *p* = .607, *d* = 0.13. There were no differences in ratings of negative/high arousal emotions between practice (*M* = 2.43 *SD* = 1.54) and no practice (*M* = 2.26, *SD* = 1.32), *F*(1, 76) = 0.28, *p* = .599, *d* = 0.13. Finally, practice (*M* = 2.73, *SD* = 1.24) and no practice groups (*M* = 2.76 *SD* = 1.10) did not differ in ratings of negative/low arousal emotions, *F*(1, 76) = 0.02, *p* = .881, *d* < 0.001. These results support prediction 7a that adding practice testing does not diminish positive ratings of a lesson.

Similarly, it was predicted that practice testing and no practice testing groups would no differ in their physiological measures of emotional arousal (prediction 7b). Those in the practice testing groups had significantly more SCRs/sec (*EMM* = 0.09, *SE* = .00) than those who did not have practice testing questions (*EMM* = 0.07, *SE* = .00), *F*(1, 74) = 12.38, *p* = .001, *d* = 0.82, after controlling for baseline measures before the lesson and movements during the lesson. They also had a higher peak-to-peak amplitude (*EMM* = 8.96, *SE* = 1.14) than the no practice testing groups (*EMM* = 5.54, *SE* = 1.14), *F*(1, 74) = 4.35, *p* = .040, *d* = 0.49. However, practice (*EMM* = 11.18, *SE* = .62) and no practice groups (*EMM* = 9.80, *SE* = .62) did not differ in their average SCL, *F*(1, 74) = 2.45, *p* = .122, *d* = 0.36. Heart rate did not differ between practice (*EMM* = 86.54, *SE* = .91) and no practice groups (*EMM* = 85.75, *SE* = .91), *F*(1, 74) = 0.38, *p* = .542, *d* = 0.14, and HRV did not differ between practice (*EMM* = 7.75, *SE* = 1.84) and no practice groups (*EMM* = 8.21, *SE* = 1.84), *F*(1, 74) = 0.03, *p* = .861, *d* < 0.001. These results do not support prediction 7b. Rather, these findings indicate that practice testing caused higher emotional arousal than no practice testing on some measures, which partially supports the alternative hypothesis that verbal practice questions may have been stressful.

### *Did Instructional Media or Practice Testing Affect Presence or Other Self-Report Measures?*

As expected, students in the IVR conditions reported significantly higher presence (*M* = 5.03, *SD* = 1.19) than those in the video conditions (*M* = 4.31, *SD* = 1.15), *F*(1, 76) = 7.75, *p* = .007, *d* = 0.64. This helps validate that IVR was successful in creating the intended immersive experience in the learners. There were no significant differences between the two media conditions or interactions involving media and practice testing on ratings of difficulty, understandability, effort, interest, enjoyment, and motivation. This suggests that desktop computers were received just as well as immersive virtual reality by learners.

There were no effects of practice testing on presence or motivation. However, there were several effects of practice testing on other self-report measures. Students who received practice testing questions (*M* = 4.59, *SD* = 1.02) rated the lesson as significantly more difficult than those who did not (*M* = 2.99, *SD* = 1.16), *F*(1, 76) = 42.82, *p* < .001, *d* = 1.50. Practice testing (*M* = 3.28, *SD* = 1.52) led participants to rate the lesson as less interesting than no practice testing (*M* = 4.21, *SD* = 1.79), *F*(1, 76) = 6.29, *p* = .014, *d* = 0.57. Practice testing (*M* = 4.56, *SD* = 1.67) made the lesson less enjoyable than no practice (*M* = 6.04, *SD* = 0.91), *F*(1, 76) = 24.00, *p* < .001, *d* = -1.12, and those who had practice testing questions (*M* = 3.51, *SD* = 1.15) reported understanding the lesson less than no practice (*M* = 4.35, *SD* = 1.27), *F*(1, 76) = 9.44, *p* = .003, *d* = -0.70, than those who did not receive practice testing questions. The practice testing groups (*M* = 4.75, *SD* = 1.20) also reported putting in more effort during the lesson than those in the non-practice testing groups (*M* = 4.14, *SD* = 1.55), *F*(1, 76) = 4.03, *p* = .048, *d* = 0.46. There were no interactions between instructional media and practice testing on presence or other self-report measures. Overall, practice testing appeared to be an aversive activity for learners, perhaps because of having to perform in front of the experimenter on a topic that was new.

### Discussion

The results from Experiment 2 revealed a similar pattern as in Experiment 1. Students performed better on transfer tests after learning from a video than learning from IVR. Although learners in IVR did not report higher extraneous cognitive load than those who viewed the video, those in the video conditions showed a higher probability of being in a high engagement cognitive state, which partially supports the notion that IVR is distracting to learning. Regarding affective processing, those in the IVR conditions reported experiencing more positive/high arousal emotions and less negative/low arousal emotions than the video conditions, which supports the prediction that IVR is more positively emotionally arousing. The IVR lesson also led to higher heart rate than the video lesson, which partially supports the hypothesis that IVR is more physiologically arousing than video. The mediation analyses revealed that positive, highly arousing emotions explained the relationship between the instructional media and retention scores.

In sum, the findings from Experiment 2 show that viewing a history lesson in a desktop video format led to higher transfer performance than viewing the same lesson in IVR, and this may be because participants who received video on a desktop computer were more cognitively engaged while viewing the video lesson and were less emotionally aroused (i.e., distracted by emotional states) than those who received the IVR lesson. The practice testing effect was not seen in this experiment. This could have been because the oral questions were stressful, and therefore were more distracting than helpful, as evidenced by the self-report ratings and physiological measures between the practice and no practice testing groups. The results of this study support the principles of the cognitive theory of multimedia learning, as supported by the hypothesis that the emotional arousal caused by IVR could be distracting.

# Chapter IV: General Discussion

## Empirical Implications

The purpose of this study was to examine the differences between learning in immersive virtual reality and more traditional lessons on a desktop computer. In two experiments, learners performed better on transfer tests after viewing a biology lesson in a PowerPoint slideshow or a history lesson in an interactive video compared to equivalent lessons in IVR. Transfer performance may be a better indicator of deep cognitive processing during learning than retention performance because it is most vulnerable to distraction during learning (Mayer & Wittrock, 1996). Distraction during learning could limit the amount of available cognitive processing capacity, so even though the learner might be able to engage in the preliminary learning process of selecting (which supports retention test performance) there might not be enough capacity left over to complete the deeper subsequent processes of organizing and integrating (which support transfer test performance).

Additionally, both experiments showed that the IVR lessons were more cognitively distracting based on EEG-based measures and more emotionally arousing based on self-report and heart-rate measures. Mediation analyses showed that measures of cognitive distraction, particularly self-reported extraneous load, and emotional distraction, particularly self-reported highly arousing positive emotions, partially explain the relationship between instructional media and learning outcomes in the form of retention tests. Although the effect of instructional media on transfer scores was more prominent than the effect on retention scores, cognitive and affective processes measured during learning did not explain the relationship between the media and transfer performance. Retention and transfer performance may differ in the underlying cognitive and affective mechanisms that predict performance. As transfer performance relies on a deeper understanding of the material than retention performance, there may be alternative underlying mechanisms or interactions between mechanisms. The relationship between instructional media and transfer performance may be more complicated to explain through the mediators measured in this study than retention performance. Overall, the results from both experiments yield converging evidence to support the distraction hypothesis in that IVR led to more cognitive and emotional distraction, which was associated with poorer performance on learning outcomes.

## Theoretical Implications

The results from both experiments can be explained by CTML. The perceptual richness of immersive virtual reality 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 lessons had extraneous sounds, animations, and interactions that were not pertinent to learning the information form 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 (Mayer et al., 1996; Mayer et al., 2001; Moreno & Mayer, 2000).

The motivational theories, such as interest theory, self-determination theory, and self-efficacy theory, were not supported in this study as increases in motivation were predicted to also increase learning outcomes. Contrary to many proponents of IVR, in both experiments, IVR groups and desktop groups did not differ in their self-reported motivation for the lesson. Although students reported enjoying the IVR lesson and being more interested in the lesson more than desktop lessons, this did not necessarily translate into motivation to learn during the lesson. IVR also increased a learner’s sense of presence, but this increased presence was also not related to his or her motivation.

## Practical Implications

A practical question is whether IVR should be used in the classroom. Based on the results of this study, it is not possible to recommend using immersive virtual reality to replace video lessons or other types of instructional media, particularly for lessons similar to the ones used in this study. Participants learned significantly more after viewing the video or PowerPoint lesson compared to viewing the IVR lesson, so the costly conversion from video to IVR may not yet be fruitful. However, some evidence has shown that IVR is useful for learning in some cases (e.g., Kohevnikov et al., 2013, Webster, 2016). Therefore, future research should determine moderating factors of the effectiveness of IVR lessons. For example, benefits of IVR could be domain specific. Additionally, specific topics within a domain could also differ in whether IVR is beneficial. One moderating factor could be 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.

Finally, because of the positive ratings by students and teachers of IVR lessons, they may be useful as pre-training tools before lessons rather than replacements for lessons. 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 principles has had robust effects on learning outcomes among other types of lessons (Mayer & Pilegard, 2014).

## Limitations and Future Directions

The practice testing effect was not replicated in this study. This could have been due to a number of reasons. First, practice testing effects may be stronger when measured with a delayed test, rather than an immediate test (Roediger & Karpike, 2006). Second, the oral format in which it was delivered did not match the final test format, which consisted of written multiple choice and short answer questions. More robust practice testing effects have been observed using paradigms in which 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 the practice test does not match the final test format (e.g., using a multiple-choice practice test and a cued recall final test; Fazio et al., 2010).

Third, asking learners to provide oral answers to oral questions may have caused more social stress than using written questions, which may have led to poorer performance on the retention and transfer tests than expected. Some parallel examples show that inducing stress and public self-consciousness (i.e., having to perform in front of an audience), specifically using the Trier Social Stress Task, reduces cognitive performance (Mohiyeddini et al., 2013). Although participants did not report different emotions felt during the lesson with and without practice testing in either experiment, there is some evidence that practice questions were stressful as participants had higher physiological arousal after practice questions in both experiments. Finally, the practice test only used 1 question per segment; other variations of practice testing have included more or longer questions, which may strengthen its effect. Future research should examine how to implement practice testing in IVR in ways that are helpful for learning.

A frequent critique of media comparison studies is that it is difficult to separate the effects of instructional media from the effects of instructional method (Clark & Feldon, 2014). For example, in this study, the use of IVR in experiment 1 was confounded with the method in which the lesson was displayed; IVR used continuous animation and spoken narration, while the PowerPoint contained static pictures and written text. In order to maximize experimental control, the same graphics and script were used to deliver equivalent academic content. However, more work is needed to experimentally control and isolate the features of IVR that caused the observed effects.

Additionally, the use of multiple measures of cognitive and emotional processing sometimes yielded inconsistent results. For example, in Experiment 2, the EDA measure of skin conductance responses per second suggest that IVR and the video lesson did not cause differences in physiological arousal. However, there were differences in heart rate, suggesting that the IVR lesson was significantly more physiologically arousing than the video lesson. This study was a first step in exploring objective measures of cognitive and emotional processing during learning with immersive media. Therefore, the results from this study may be insightful in determining which physiological measures were most sensitive to the types of lessons and equipment that were used in the study. For example, EDA measures may be more sensitive to detecting sudden changes in physiological arousal than heart rate, particularly aversive stimuli, rather than responses to prolonged exposure to a stimuli, which is evidenced by the differences in EDA measures between practice and no/practice groups in both experiments and no differences between the IVR and desktop lesson. Future research should determine best practices for acquiring physiological measurements during multimedia lessons.

Finally, it is possible that the results of the study may be explained by a *novelty effect* of IVR in which the use of the new technology in itself caused poorer learning outcomes and/or distraction. Many participants reported that it was their first time using an IVR console; wearing the headset, holding the controllers, and moving the controllers in a specific way could have all been sources of distraction. For example, the unfamiliarity of using the IVR interface could have induced extraneous processing until the user had learned, and was comfortable with, the interactions with the interface. Additional cognitive processing would also have been required to ignore the feeling of wearing a headset or monitor the position of the controllers. These added sources of processing could have taken up some of the cognitive resources that should have been used for essential or generative processing of the incoming information. Future research could examine this novelty effect by comparing learning outcomes from habitual VR users to novice users, or using a training intervention to familiarize learners with the mechanisms required to interact with the IVR interface. It could be possible that as IVR proliferates classrooms and households in the coming years, this novelty effect would be diminished and learning from IVR would be equivalent to learning from other multimedia technologies.

# References

Anderson, C.W., & Bratman, J.A. (2008). Translating thoughts into actions by finding patterns in brainwave. *Proceedings of the Fourteenth Yale Workshop on Adaptive and Learning Systems* (pp. 1–6). Yale University, New Haven, CT.

Andreassi, J. L. (2010). Electrodermal activity and behavior. *Psychophysiology: human behavior and physiological response* (5th ed., pp. 314-340). Lawrence Erlbaum Associates, Publishers. https://doi.org/10.4324/9780203880340

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review, 22* (4), 425-438. http://doi.org/10.1007/s10648-010-9130-y

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*, *Volume 2* (pp. 89–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Australian Broadcasting Company. (2018). Kokoda VR [Computer Application]. https://www.abc.net.au/btn/kokodaVR/

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

Bailenson, J., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences*, *17*(1), 102-141. https://doi.org/10.1080/10508400701793141

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215. http://doi.org/10.1037/0033-295X.84.2.191

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

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

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., & 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., Olmstead, R. E., Tremoulet, P. D., & 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), B231-244.

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 University Press. https://doi.org/10.1017/CBO9780511546396

Brown, P. C., Rodiger, H. L. III., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* 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

Chemers, M. M., Hu, L., & Garcia, B. F. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology, 93*(1), 55–64. https://doi.org/10.1037//0022-0663.93.1.55

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-459. https://doi.org/10.2307/1170217

Clark, R. E. (1994). Media will never influence learning. *Educational Technology, Research and Development, 42*(2), 21-29. https://doi.org/10.1007/BF02299088

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). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.009

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers. https://doi.org/10.4324/9780203771587

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

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. Cacioppo, L. G. Tassinary, & G. G. Bernston (Eds.), *The handbook of* *psychophysiology* (3rd ed., pp. 159-181). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*(1), 105-115. https://doi.org/10.1037/h0030644

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.* Cambridge University Press. https://doi.org/10.1017/CBO9781107707085

Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 220-227. https://doi.org/10.1177/2372732216655542

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest.

*Journal of Educational Psychology, 89*(1), 92–102. https://doi.org/10.1037/0022-0663.89.1.92

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414–434. https://doi.org/10.1037/0022-0663.90.3.414

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* 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

Izard, C. E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology, 60*, 1–25. https://doi.org/10.1146/annurev.psych.60.110707.163539

Kafai,Y. B. (2006). Playing and making games for learning: Instructionist and constructionist perspectives for game studies. *Games and Culture, 1*(1), 36–40. https://doi.org/10.1177/1555412005281767

Klimesch, W., Schack, B., & Sauseng, P. (2005). The functional significance of theta and upper alpha oscillations for working memory. *Experimental Psychology, 52*(2), 99–108. https://doi.org/10.1027/1618-3169.52.2.99

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

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

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

Kozma, R. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development, 42*(2), 7-19. https://doi.org/10.1007/BF02299087

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), 261–73. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

LaValle, S. M. (2019). *Virtual Reality*. Cambridge University Press.

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and Brain Sciences, 28*, 169–194. https://doi.org/10.1017/S0140525X0500004X

Ligorio, M. B., & van Veen, K. (2006). Constructing a successful cross-national virtual learning environment in primary and secondary education. *Association for the Advancement of Computing in Education Journal, 14*(2), 103–128.

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. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, *60*, 225-236. https://doi.org/10.1016/j.learninstruc.2017.12.007

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. (2002). Multimedia learning. *Psychology of Learning and Motivation*, *41,* 85-139. https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). 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). 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. https://doi.org/10.1016/j.learninstruc.2014.02.004

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). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.009

Mayer, R. E., Griffith, E., Naftaly, I., & Rothman, D. (2008). Increased interestingness of extraneous details leads to decreased learning. *Journal of Experimental Psychology: Applied*, *14*(4), 329 –339. https://doi.org/10.1037/a0013835

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). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.009

Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In D.C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47–62). Prentice Hall International.

McMahan, A. (2003). Immersion, engagement, and presence: A method for analyzing 3-D video games. In W. Mark & B. Perron (Eds.), *The video game theory reader* (pp. 67–86). Routledge.

Mohiyeddini, C., Bauer, S., & Semple, S. (2013). Public self-consciousness moderates the link between displacement behaviour and experience of stress in women. *Stress: The International Journal on the Biology of Stress, 16*(4), 384-392. https://doi.org/10.3109/10253890.2012.755171

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

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. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

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), 108-34. 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, 1,* 133-141. Bucharest, Romania. https://doi.org/10.12753/2066-026X-15-020

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*(1), 27-45. https://doi.org/10.1007/s10648-011-9179-2

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning (2nd ed.*, pp.

27-42). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.009

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

Pekrun, R. (2014). *Emotions and learning* (Educational Practices Series, Vol. 24). International Academy of Education (IAE) and International Bureau of Education (IBE) of the United Nations Educational, Scientific and Cultural Organization (UNESCO).

Pekrun, R. (2017). Achievement emotions. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (pp. 251–271). Guilford Press.

Pietsch, J., Walker, R., & Chapman, E. (2003). The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of Educational Psychology, 95*(3), 589-603. https://doi.org/10.1037/0022-0663.95.3.589

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

Plass, J. L., & Kaplan, U. (2015). Emotional design in digital media for learning. In S. Tettegah, & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 131-162). New York: Elsevier. https://doi.org/10.1016/B978-0-12-801856-9.00007-4

Poythress, M. L., Russell, C., Siegel, S. I., Tremoulet, P. D., Craven, P. Berka, C., Levendowski, D. J., Chang, D., Baskin, A., Champney, R.K., Hale, K., & Milham, L. (2006). Correlation between expected workload and EEG indices of cognitive workload and task engagement. In Schmorrow, D., Stanney, K., & Reeves, L. (Eds.), *Augmented Cognition: Past, Present and Future* (pp. 32-44). Strategic Analysis, Inc.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology, 14*(2), 154-166. http://doi.org/10.1037/a0019440

Renninger, K. A., & Hidi, S. (2016). *The power of interest for motivation and engagement*. Routledge. https://doi.org/10.4324/9781315771045

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

Ryan, M. L. (2001). *Narrative as virtual reality: Immersion and interactivity in literature and electronic media.* Johns Hopkins University Press.

Ryan, M. L. (2008). Interactive narrative, plot types, and interpersonal relations. In U. Spierling & N. Szilas (Eds.), *Interactive storytelling* (pp. 6–13). Springer-Verlag. https://doi.org/10.1007/978-3-540-89454-4_2

Ryan, R. M., & Deci, E. E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78. https://doi.org/10.1037110003-066X.55.1.68

Ryan, R. M., Rigby, C. S., & Przybylski, A. K. (2006). Motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*, 347-365. https://doi.org/10.1007/s11031-006-9051-8

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

Salomon, G. (1979). *Interaction of media, cognition, and learning.* Lawrence Erlbaum Associates, Inc., Publishers

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

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

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (1st ed., pp. 197–223). Routledge.

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). Erlbaum.

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*(3-4), 207-231. https://doi.org/10.1080/00461520.1991.9653133

Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school* (2nd ed., pp. 34–54). Routledge.

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influences on children's self-efficacy and achievement. *Journal of Educational Psychology, 77*(3), 313–322. https://doi.org/10.1037/0022-0663.77.3.313

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

Slater, M. (2003). A note on presence terminology. *Presence-Connect, 3*(3), 1-5.

Sourin, A., Sourina, O. & Prasolova-Førland, E. (2006). Cyber-learning in cyberworlds. *Journal of Cases on Information Technology, 8*(4), 55–70. https://doi.org/10.4018/jcit.2006100105

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*. 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]. https://thebodyvr.com/journey-inside-a-cell/

Tucker, D. M. (2007). *Mind from body: Experience from neural structure.* Oxford University Press.

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

van Dooren, M., de Vries, J. J. G., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior, 106*(2), 298-304. https://doi.org/10.1016/j.physbeh.2012.01.020

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

Wentzel, K. & Miele, D. B.. (Eds.). (2016) *Handbook of motivation at school* (2nd ed). Routledge.

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F. R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., & Jäncke, P. (2007). A process model of the formation of spatial presence experiences. *Media Psychology, 9*, 493–525. https://doi.org/10.1080/15213260701283079

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

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*(4), 345-376. https://doi.org/10.1207/s15326985ep2404_2

Yee, N., & Bailenson, J. N. (2007). The proteus effect: Self transformations in virtual reality. *Human Communication Research, 33*, 271–290. https://doi.org/10.1111/j.1468-2958.2007.00299.x

## Tables and Figures

Table 1

*Dependent Measures for Experiments 1 and 2*

| Measure | Definition |
|---|---|
| **Learning outcomes** | |
| Retention | Number of correct responses out of 16 multiple choice questions |
| Transfer | Number of correct statements in 4 open-ended questions (15 points total in Experiment 1; 17 points total in Experiment 2) |
| **Cognitive processing** | |
| Self-report | |
| Extraneous Load | Average score between 1 (strongly disagree) and 7 (strongly agree) on a 3-item subscale. Sample item: “I felt distracted during the lesson.” |
| Essential Load | Average score between 1 (strongly disagree) and 7 (strongly agree) on a 3-item subscale. Sample item: “I was trying to learn the main facts from the lesson.” |
| Generative Load | Average score between 1 (strongly disagree) and 7 (strongly agree) on a 3-item subscale Sample item: “I was trying to make connections between the material and things I already know.” |
| EEG | |
| High Engagement | Average probability that the learner’s cognitive state is in a high engagement state during the lesson |
| Low Engagement | Average probability that the learner’s cognitive state is in a low engagement state during the lesson |
| Distraction | Average probability that the learner’s cognitive state is in a distraction state during the lesson |
| Sleep Onset | Average probability that the learner’s cognitive state is in a sleep onset state during the lesson |
| Workload | Learner’s average cognitive workload on a scale from 0 to 1 during the lesson; higher values indicate more workload. 0- |

| | |
|---|---|
| | .39 indicates boredom, .40-.69 indicates ideal workload, and .7-1 indicates cognitive workload |
| **Affective processing** | |
| Self-report | |
| Positive, high arousal | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “I felt excited.” |
| Positive, low arousal | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “I felt calm.” |
| Negative, high arousal | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “I felt frustrated.” |
| Negative, low arousal | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “I felt bored.” |
| EDA | |
| Skin conductance level | Average skin conductance in microsiemens from the tonic electrodermal activity signal during the lesson |
| Skin conductance responses/sec | Average number of non-specific skin conductance responses divided by the length of the lesson in seconds |
| Peak-to-peak amplitude | The difference between the lowest and highest skin conductance level from the phasic electrodermal activity signal during the lesson |
| ECG | |
| Heart rate | Average heart rate during the lesson (BPM) |
| Heart rate variability (HRV) | Average variation in the time interval between heartbeats during the lesson in the high frequency range (0.15 - 0.40Hz, sec$_2$) |
| **Self-report items** | |
| Presence | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 7-item subscale. Sample item: “I felt like I was really in the environment.” |
| Difficulty | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “The lesson was difficult.” |

| | |
|---|---|
| Understanding | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “I have a good understanding of the material.” |
| Effort | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “I put a lot of effort in the lesson.” |
| Interest | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “I was interested in this topic before the experiment.” |
| Enjoyment | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “I enjoyed learning this way.” |
| Motivation | Average score between 1 (strongly disagree) to 7 (strongly agree) on a 2-item subscale. Sample item: “I felt motivated to understand the lesson.” |

Table 2

*Means (SD) of Learning Outcome Scores for All conditions in Experiments 1 and 2*

| | **Retention score** | | **Transfer score** | |
|---|---|---|---|---|
| | IVR | Desktop | IVR | Desktop |
| **Experiment 1** | | | | |
| 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) |
| **Experiment 2** | | | | |
| No practice | 8.30 (2.58) | 9.40 (2.95) | 2.85 (1.75) | 4.10 (2.55) |
| Practice | 10.00 (2.68) | 9.65 (2.13) | 3.28 (2.34) | 4.08 (2.18) |

*Note.* The desktop condition was a PowerPoint slideshow in Experiment 1 and a video in Experiment 2. The retention score for both experiments was out of a total of 16 points. For Experiment 1, transfer score was out of 15 possible points, and for Experiment 2, transfer score was out of 17 possible points.

Table 3

*Means (SD) of Self-Reported Cognitive Load Scores During the Lesson for All Conditions in Experiments 1 and 2*

| | **Extraneous load** | | **Essential load** | | **Generative load** | |
|---|---|---|---|---|---|---|
| **Exp. 1** | 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) |
| **Exp. 2** | IVR | Video | IVR | Video | IVR | Video |
| No practice | 3.55 (1.88) | 2.73 (1.34) | 5.07 (1.49) | 5.47 (0.91) | 4.98 (1.21) | 4.73 (0.82) |
| Practice | 4.02 (1.44) | 3.67 (1.14) | 5.65 (0.83) | 5.47 (0.96) | 5.15 (0.94) | 4.91(1.09) |

*Note.* Each averaged score was on a scale from 1-7, with higher scores indicating increased cognitive load.

Table 4

*Means (SD) of EEG-based Cognitive Metrics During the Lesson for All Conditions in Experiments 1 and 2*

| | **High Engagement** | | **Low Engagement** | | **Distraction** | | **Sleep Onset** | | **Workload** | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Exp. 1** | 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) |
| **Exp. 2** | IVR | Video | IVR | Video | IVR | Video | IVR | Video | IVR | Video |
| No Practice | .36 (.17) | .35 (.19) | .40 (.15) | .37 (.15) | .10 (.10) | .13 (.19) | .09 (.11) | .10 (.09) | .44 (.20) | .44 (.15) |
| Practice | .41 (.20) | .41 (.18) | .41 (.23) | .38 (.14) | .08 (.12) | .12 (.17) | .08 (.11) | .07 (.07) | .44 (.19) | .41 (.12) |

*Note.* The high engagement, low engagement, distraction, and sleep onset cognitive engagement metrics were on a scale from 0-1, with higher scores indicating a greater probability of the learner being in that cognitive state. The workload metric was on a scale from 0-1, with 0-.39 indicating boredom, .40-.69 indicating ideal cognitive workload, and .70-1.00 indicating cognitive overload.

Table 5

*Frequencies of Type of Cognitive Workload During the Lesson for All Conditions in Experiments 1 and 2*

| | | Cognitive Workload Type | | |
|---|---|---|---|---|
| | | Bored | Ideal | Overload |
| **Experiment 1** | | | | |
| IVR | No Practice | 9 | 6 | 0 |
| | Practice | 3 | 9 | 3 |
| PPT | No Practice | 2 | 13 | 1 |
| | Practice | 2 | 12 | 1 |
| **Experiment 2** | | | | |
| IVR | No Practice | 6 | 12 | 2 |
| | Practice | 9 | 11 | 0 |
| PPT | No Practice | 8 | 12 | 0 |
| | Practice | 12 | 8 | 0 |

Table 6

*Means (SD) of Self-Reported Emotional States During the Lesson for All Conditions in Experiments 1 and 2*

| | **Positive, High Arousal** | | **Positive, Low Arousal** | | **Negative, High Arousal** | | **Negative Low Arousal** | |
|---|---|---|---|---|---|---|---|---|
| **Exp. 1** | 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) |
| **Exp. 2** | IVR | Video | IVR | Video | IVR | Video | IVR | Video |
| No Practice | 4.63 (1.67) | 3.63 (1.38) | 4.45 (1.23) | 4.55 (1.01) | 1.73 (1.03) | 2.80 (1.39) | 2.28 (1.07) | 3.25 (0.92) |
| Practice | 4.83 (1.17) | 3.53 (1.39) | 4.55 (1.31) | 4.73 (1.18) | 2.05 (1.33) | 2.80 (1.68) | 2.43 (1.23) | 3.03 (1.21) |

*Note.* Each averaged score was on a scale from 1-7, with higher scores indicating increased emotion.

Table 7

*Means (SD) of Electrodermal Activity Measures During the Lesson for All Conditions in Experiments 1 and 2*

| | **SCL** | | **SCR/sec** | | **Peak-to-peak** | |
|---|---|---|---|---|---|---|
| **Experiment 1** | 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) |
| **Experiment 2** | IVR | Video | IVR | Video | IVR | Video |
| No practice | 10.54 (8.49) | 8.78 (6.84) | .08 (.03) | .05 (.04) | 8.64 (12.57) | 3.49 (3.61) |
| Practice | 13.04 (7.07) | 9.58 (5.13) | .10 (.03) | .08 (.04) | 7.92 (10.31) | 8.96 (5.59) |

*Note*. EDA was measured in microsiemens. SCL = average skin conductance level from the tonic signal. SCR/sec = non-specific skin conductance responses per second. Peak-to-peak = difference between highest and lowest points of phasic EDA signal.

Table 8

*Means (SD) of Electrocardiography Measures During the Lesson for All Conditions in Experiments 1 and 2*

| | **Heart rate** | | **Heart rate variability** | |
|---|---|---|---|---|
| **Experiment 1** | 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) |
| **Experiment 2** | IVR | Video | IVR | Video |
| No Practice | 90.32 (10.95) | 80.49 (11.24) | 8.56 (17.99) | 2.75 (2.90) |
| Practice | 90.88 (10.31) | 82.87 (13.52) | 12.07 (20.98) | 8.53 (17.49) |

*Note.* Heart rate was measured in beats per minute (BPM). Heart rate variability (HRV) was measured using the high-frequency band (0.15 - 0.40Hz, ms$_2$).

Table 9

*Cognitive and Affective Predictors of Retention Score in Experiment 1*

| | B | SE | β | F | $R_2$ |
|---|---|---|---|---|---|
| Self-report cognitive load | | | | 4.12** | .18 |
| Extraneous load | -.86 | .28 | -.39** | | |
| Essential load | .40 | .47 | .13 | | |
| Generative load | -.11 | .45 | -.04 | | |
| EEG-based cognitive metrics | | | | 1.04 | .09 |
| High Engagement | -1.56 | 5.18 | -.07 | | |
| Low Engagement | -6.80 | 5.21 | -.40 | | |
| Distraction | -8.66 | 6.07 | -.28 | | |
| Sleep Onset | -4.80 | 5.64 | -.17 | | |
| Workload | 3.47 | 3.06 | .16 | | |
| Self-report emotional states | | | | 0.52 | .04 |
| Positive valence, high arousal | -.17 | .32 | -.08 | | |
| Positive valence, low arousal | .26 | .44 | -.09 | | |
| Negative valence, high arousal | .19 | .90 | .03 | | |
| Negative valence, low arousal | -.69 | .55 | -.20 | | |
| EDA and ECG metrics | | | | 1.47 | .12 |
| Skin conductance level | .10 | .09 | .18 | | |
| Skin conductance responses/sec | -40.85 | 15.31 | -.40* | | |
| Peak-to-peak amplitude change | .03 | .15 | .03 | | |
| Heart rate | .001 | .04 | .004 | | |
| High-frequency heart rate variability | .001 | .01 | .02 | | |

*Note.* * *p* < .05, ** *p* < .01, *** *p* < .001

Table 10

*Cognitive and Affective Predictors of Transfer Score in Experiment 1*

| | B | SE | β | F | $R^2$ |
|---|---|---|---|---|---|
| Self-report cognitive load | | | | 3.37* | .15 |
| Extraneous load | -.35 | .14 | -.30* | | |
| Essential load | .43 | .24 | .26 | | |
| Generative load | -.46 | .23 | -.30 | | |
| EEG-based cognitive metrics | | | | 2.93* | .21 |
| High Engagement | -3.89 | 2.48 | -.36 | | |
| Low Engagement | -7.50 | 2.50 | -.85** | | |
| Distraction | -7.59 | 2.91 | -.48* | | |
| Sleep Onset | -8.13 | 2.71 | -.55** | | |
| Workload | 2.26 | 1.47 | .21 | | |
| Self-report emotional states | | | | 1.31 | .09 |
| Positive valence, high arousal | -.21 | .16 | -.20 | | |
| Positive valence, low arousal | .40 | .22 | .25 | | |
| Negative valence, high arousal | -.01 | .45 | -.01 | | |
| Negative valence, low arousal | .00 | .28 | .00 | | |
| EDA and ECG metrics | | | | .93 | .08 |
| Skin conductance level | .03 | .05 | .10 | | |
| Skin conductance responses/sec | -11.25 | 8.06 | -.22 | | |
| Peak-to-peak amplitude change | .05 | .08 | .11 | | |
| Heart rate | -.02 | .02 | -.09 | | |
| High-frequency heart rate variability | -.01 | .01 | -.13* | | |

*Note*. * *p* < .05, ** *p* < .01, *** *p* < .001

Table 11

*Cognitive and Affective Predictors of Retention Score in Experiment 2*

| | B | SE | β | F | $R^2$ |
|---|---|---|---|---|---|
| Self-report cognitive load | | | | 0.91 | .04 |
| Extraneous load | -.16 | .20 | -.09 | | |
| Essential load | .46 | .32 | .18 | | |
| Generative load | -.27 | .34 | -.10 | | |
| EEG-based cognitive metrics | | | | 0.57 | .04 |
| High Engagement | .16 | 3.54 | .01 | | |
| Low Engagement | -1.78 | 3.72 | -.11 | | |
| Distraction | -3.08 | 3.95 | -.17 | | |
| Sleep Onset | -2.57 | 5.22 | -.10 | | |
| Workload | 1.41 | 1.90 | .09 | | |
| Self-report emotional states | | | | 1.43 | .07 |
| Positive valence, high arousal | .36 | .21 | .21 | | |
| Positive valence, low arousal | .40 | .27 | .18 | | |
| Negative valence, high arousal | .05 | .25 | .03 | | |
| Negative valence, low arousal | .20 | .31 | .09 | | |
| EDA and ECG metrics | | | | 0.54 | .04 |
| Skin conductance level | -.03 | .06 | -.09 | | |
| Skin conductance responses/sec | 8.35 | 10.45 | .12 | | |
| Peak-to-peak amplitude change | -.02 | .04 | -.07 | | |
| Heart rate | .02 | .03 | .08 | | |
| High-frequency HRV | -.02 | .02 | -.14 | | |

*Note*. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 12

*Cognitive and Affective Predictors of Transfer Score in Experiment 2*

| | B | SE | β | F | $R^2$ |
|---|---|---|---|---|---|
| Self-report cognitive load | | | | 2.05 | .08 |
| Extraneous load | -.34 | .16 | -.23* | | |
| Essential load | .32 | .27 | .15 | | |
| Generative load | -.30 | .29 | -.13 | | |
| EEG-based cognitive metrics | | | | 0.52 | .03 |
| High Engagement | .22 | 3.03 | .02 | | |
| Low Engagement | -2.50 | 3.18 | -.19 | | |
| Distraction | -1.67 | 3.38 | -.11 | | |
| Sleep Onset | -1.07 | 4.47 | -.05 | | |
| Workload | .28 | 1.63 | .02 | | |
| Self-report emotional states | | | | 2.59* | .12 |
| Positive valence, high arousal | .09 | .18 | .06 | | |
| Positive valence, low arousal | .51 | .22 | .26* | | |
| Negative valence, high arousal | -.15 | .21 | -.09 | | |
| Negative valence, low arousal | .57 | .26 | .30* | | |
| EDA and ECG metrics | | | | 1.18 | .07 |
| Skin conductance level | -.05 | .05 | -.16 | | |
| Skin conductance responses/sec | -1.63 | 8.75 | -.03 | | |
| Peak-to-peak amplitude change | .01 | .04 | .03 | | |
| Heart rate | .02 | .02 | .13 | | |
| High-frequency HRV | -.03 | .02 | -.23 | | |

*Note.* * $p < .05$, ** $p < .01$, *** $p < .001$

Table 13

*Means (SD) of Post-Lesson Self-Report Scales for All Conditions in Experiment 1*

| | IVR | PPT |
|---|---|---|
| Presence | | |
| No Practice | 5.18 (0.71) | 3.71 (0.79) |
| Practice | 5.06 (0.90) | 3.42 (1.21) |
| Difficulty | | |
| No Practice | 2.77 (0.98) | 3.50 (1.25) |
| Practice | 3.00 (1.20) | 3.73 (1.40) |
| Understandability | | |
| No Practice | 4.60 (0.97) | 4.56 (1.29) |
| Practice | 4.53 (1.03) | 4.17 (1.60) |
| Effort | | |
| No Practice | 4.37 (1.54) | 3.88 (1.47) |
| Practice | 4.10 (1.14) | 4.27 (1.59) |
| Interest | | |
| No Practice | 4.83 (1.21) | 4.22 (1.46) |
| Practice | 4.67 (1.59) | 3.73 (1.94) |
| Enjoyment | | |
| No Practice | 6.13 (1.11) | 4.16 (1.34) |
| Practice | 5.53 (1.87) | 3.80 (1.93) |
| Motivation | | |
| No Practice | 4.93 (1.50) | 4.56 (1.22) |
| Practice | 5.00 (1.39) | 4.10 (1.91) |

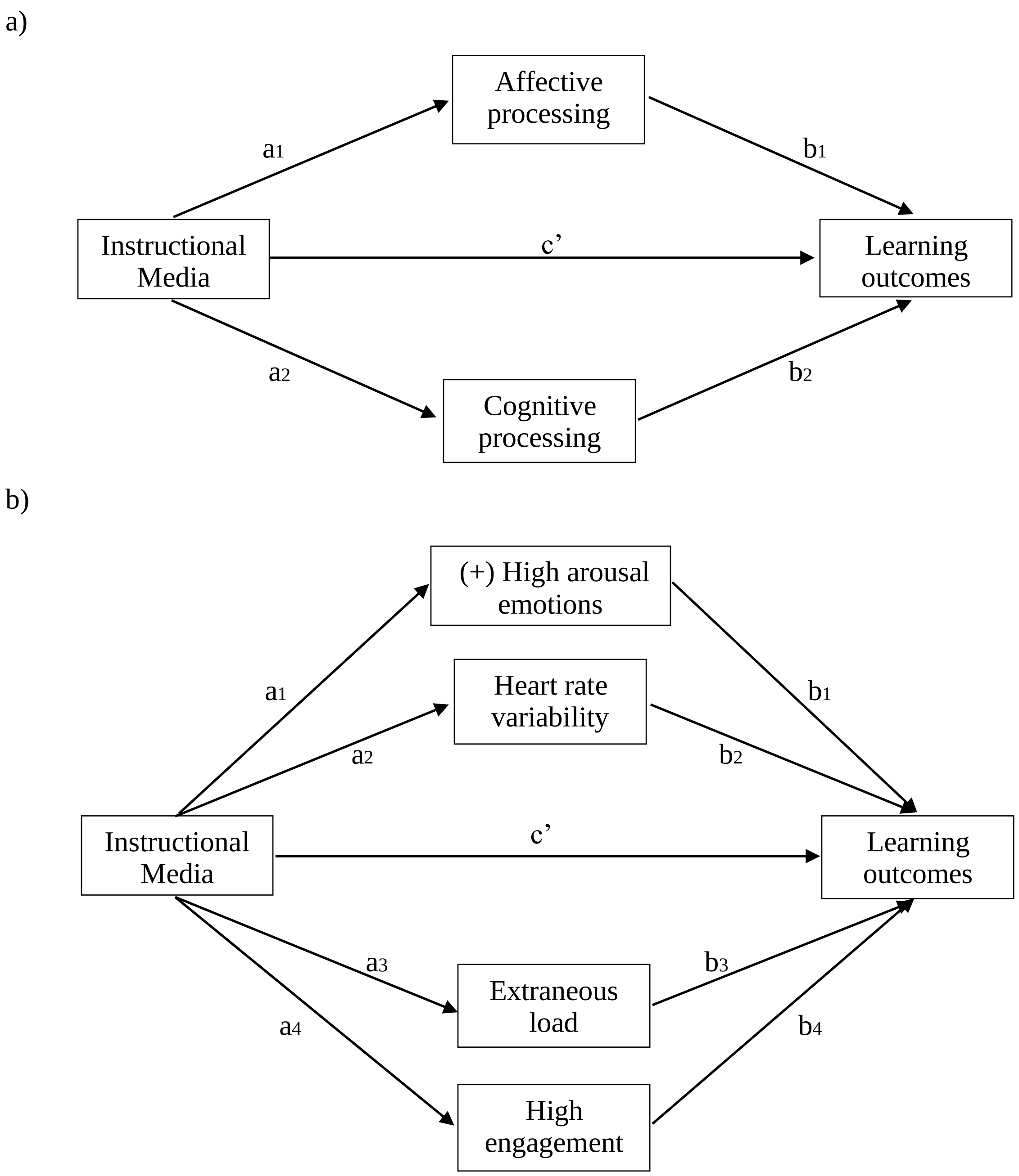
Table 14

*Means (SD) of Post-lesson Self-Report Scales for All Conditions in Experiment 2*

| | IVR | Video |
|---|---|---|
| Presence | | |
| No Practice | 4.90 (1.28) | 4.57 (1.12) |
| Practice | 5.16 (1.10) | 4.04 (1.15) |
| Difficulty | | |
| No Practice | 3.13 (1.39) | 2.85 (0.89) |
| Practice | 4.75 (1.07) | 4.43 (0.96) |
| Understandability | | |
| No Practice | 4.18 (1.32) | 4.53 (1.22) |
| Practice | 3.45 (1.25) | 3.58 (1.08) |
| Effort | | |
| No Practice | 3.70 (1.71) | 4.58 (1.26) |
| Practice | 4.85 (1.42) | 4.65 (0.95) |
| Interest | | |
| No Practice | 4.53 (2.01) | 3.90 (1.54) |
| Practice | 3.30 (1.65) | 3.25 (1.44) |
| Enjoyment | | |
| No Practice | 6.28 (0.80) | 5.80 (0.97) |
| Practice | 4.43 (1.72) | 4.70 (1.65) |
| Motivation | | |
| No Practice | 4.78 (1.42) | 4.90 (1.07) |
| Practice | 4.30 (1.50) | 4.38 (1.07) |

**Figure 1**

*Proposed Cognitive-Affective Distraction Model of Virtual Reality Learning Environments*

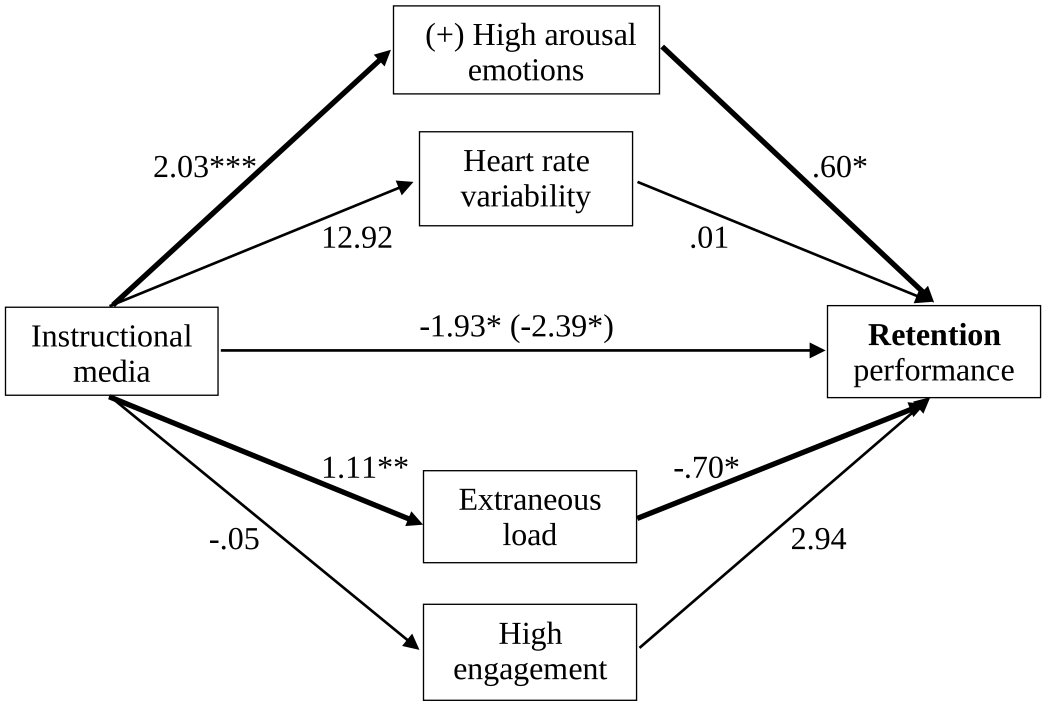


*Note.* These models show a general (1a) and specific (1b) proposed cognitive-affective distraction models of immersive virtual reality learning environments, such that instructional media cause differences in cognitive and affective processing, which cause increases or decreases in learning outcomes.

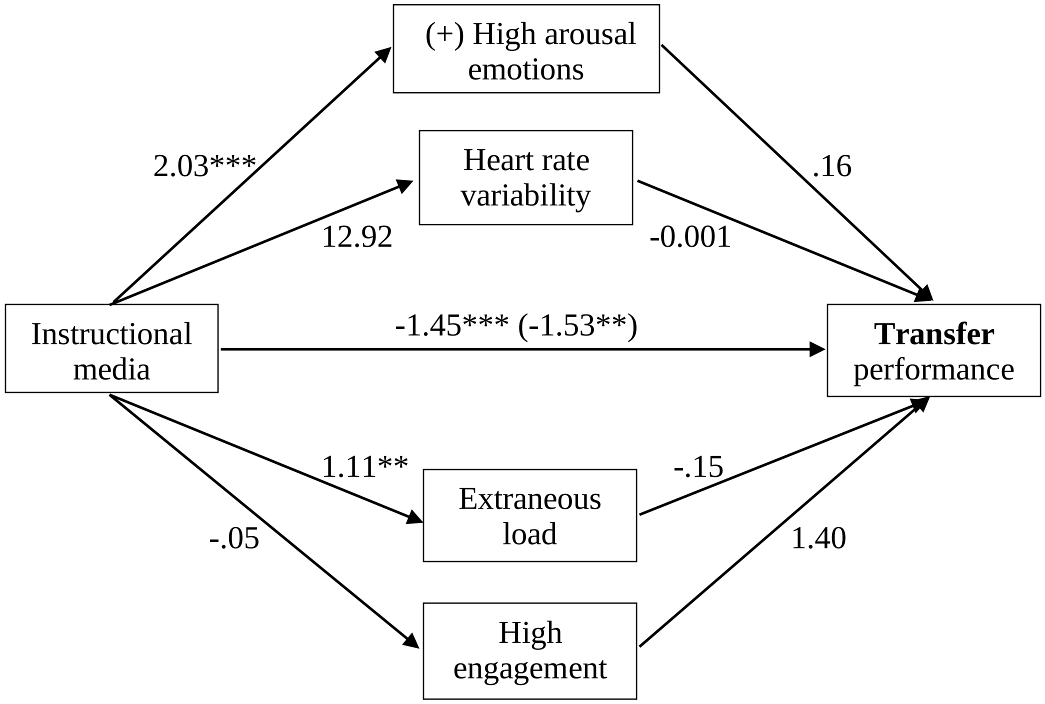
**Figure 2**

*Experiment 1 Mediation Models Predicting Learning Outcomes*

a)



b)

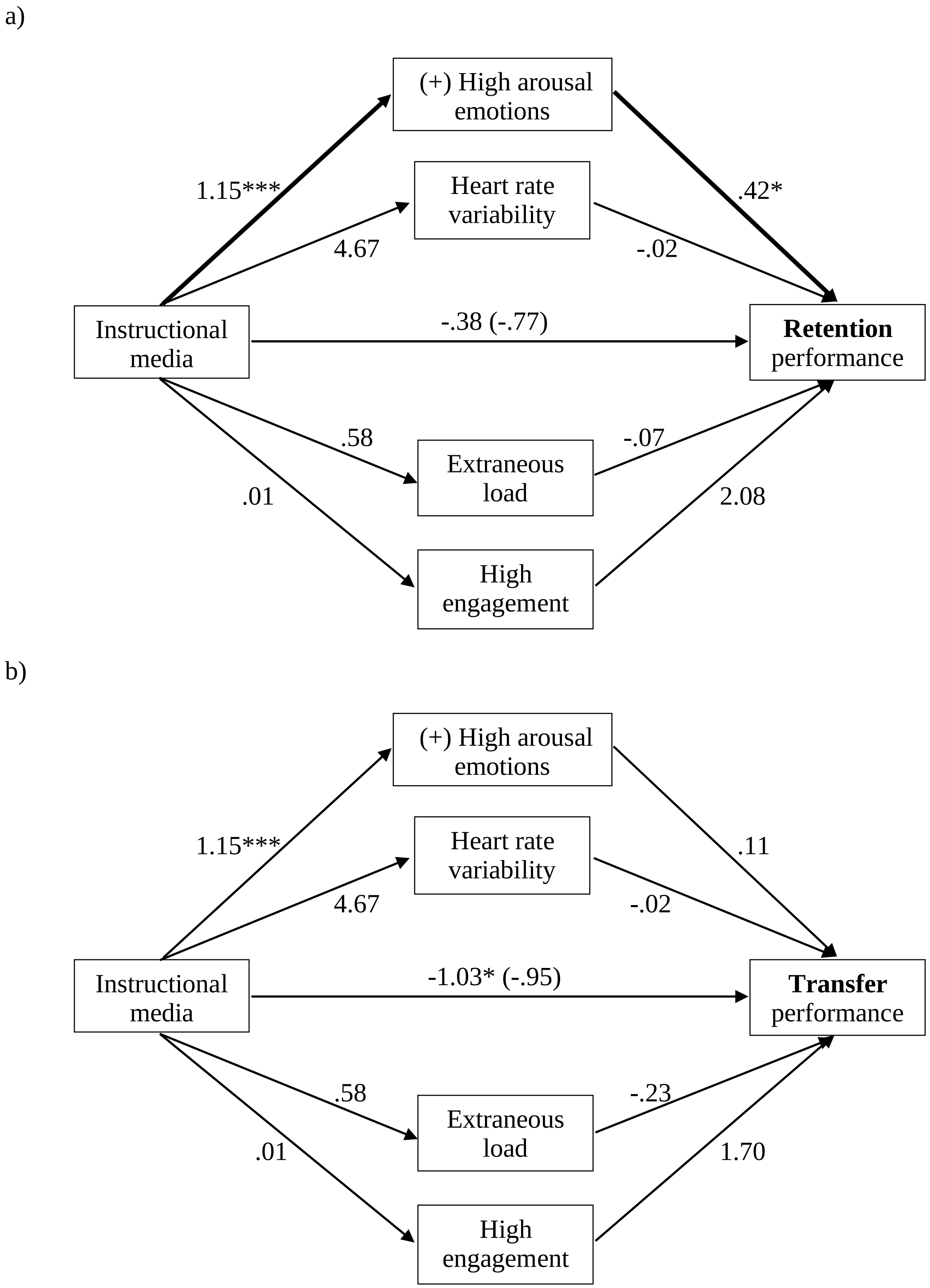


*Note*. These models explain the relationship between instructional media and (a) retention performance and (b) transfer performance for a science lesson. Instructional media was coded as a categorial variable with the PPT condition as a 0 and the IVR condition as a 1. Coefficients represent unstandardized 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).

$*p < .05$, $**p < .01$, $***p < .001$

**Figure 3**

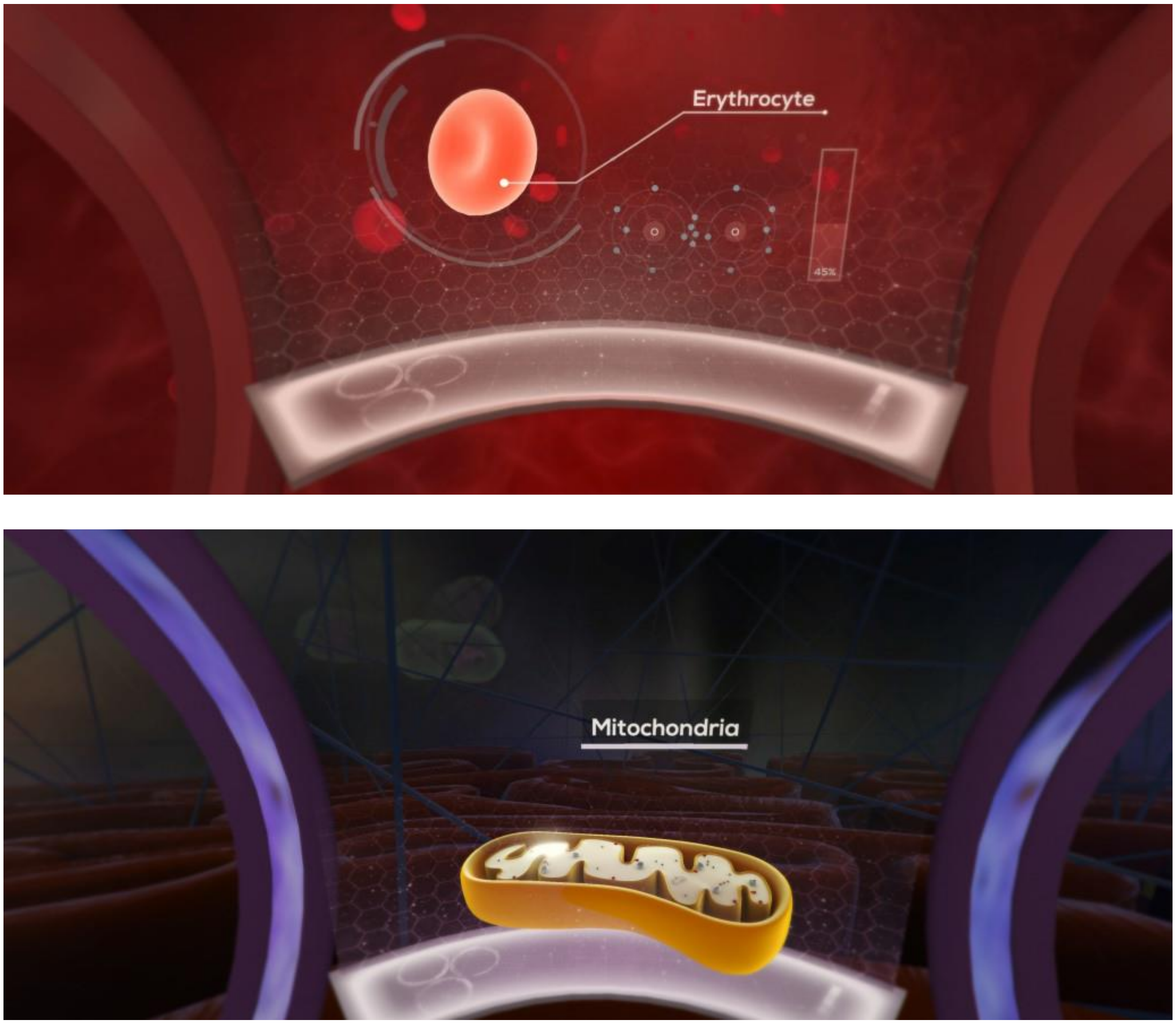
*Experiment 2 Mediation Models Predicting Learning Outcomes*



*Note*. These models explain the relationship between instructional media and (a) retention performance and (b) transfer performance for a history lesson. Instructional media was coded as a categorial variable with the video 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 performance after controlling for the mediators. Bold lines indicate significant indirect effects (i.e., mediation). $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

**Figure 4**

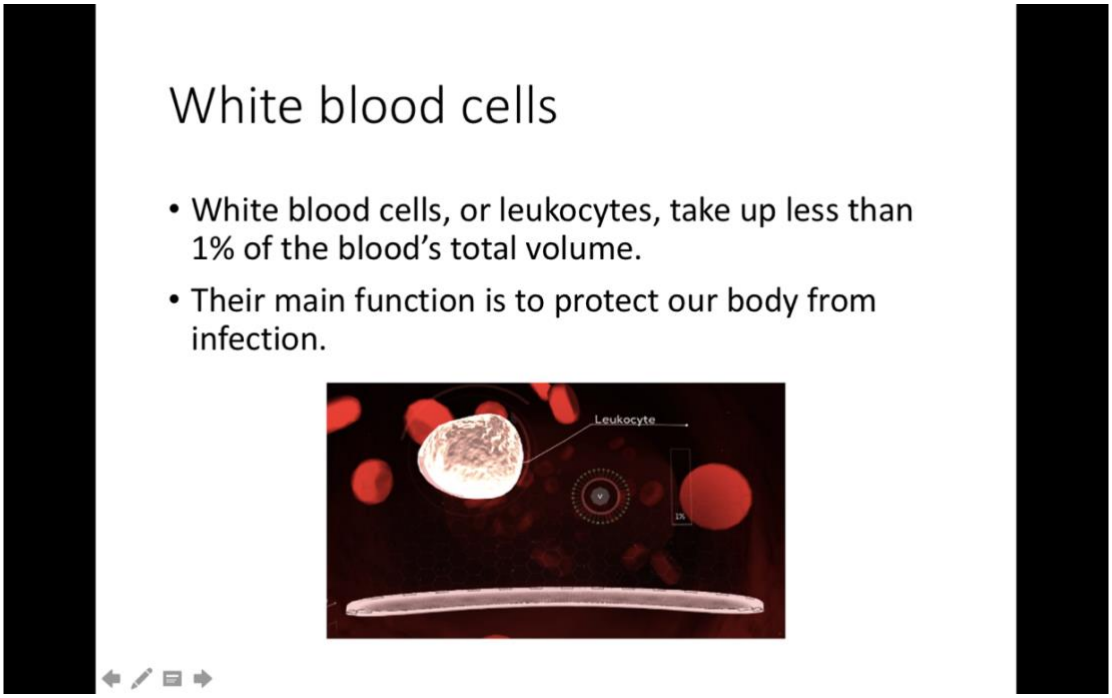
*The Body VR: Journey Inside a Cell in the IVR Conditions*



*Note*. The images show screenshots from the lesson displayed in an HTC Vive head-mounted display in Experiment 1.

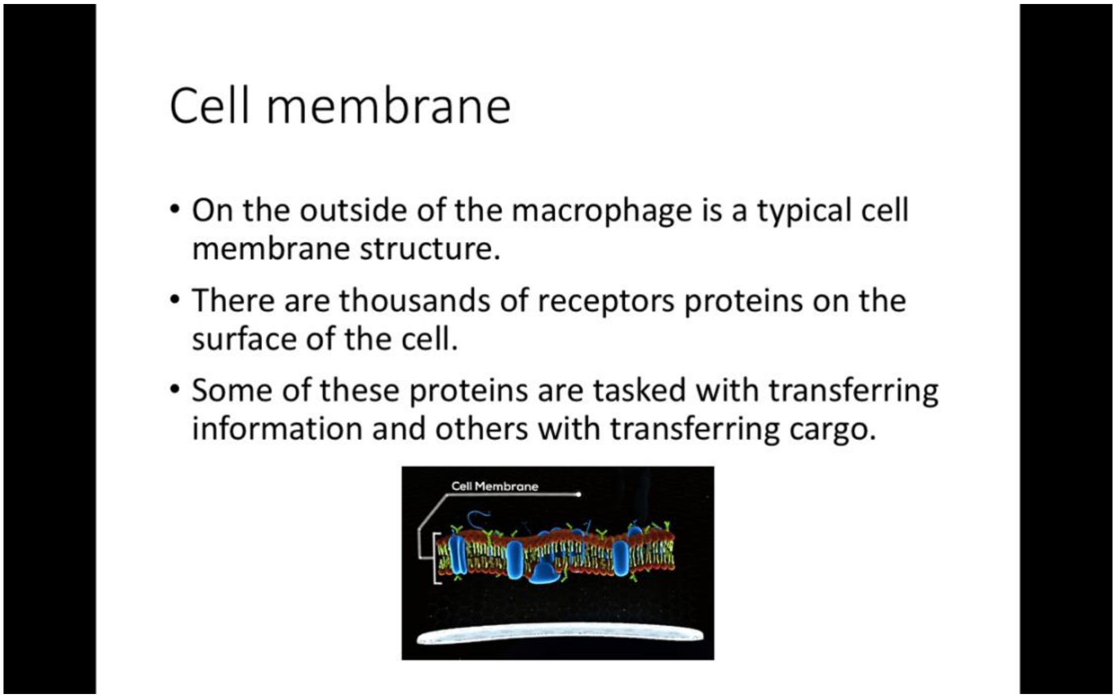
**Figure 5**

*The Body VR: Journey Inside a Cell in the Desktop Conditions*



## White blood cells

- White blood cells, or leukocytes, take up less than 1% of the blood’s total volume.
- Their main function is to protect our body from infection.



## Cell membrane

- On the outside of the macrophage is a typical cell membrane structure.
- There are thousands of receptors proteins on the surface of the cell.
- Some of these proteins are tasked with transferring information and others with transferring cargo.

*Note*. The images show screenshots from the PowerPoint-style lesson adapted from *The Body VR: Journey into a Cell* in Experiment 1.

**Figure 6**

*Kokoda VR*

*Note.* The images show screenshots from *Kokoda VR* displayed in a 360-degree YouTube video in Experiment 2. The graphics in the IVR lesson were the same, but displayed with the HTC Vive head-mounted display.

# **Appendices**

## Appendix A

### Experiment 1: Pre-lesson Questionnaire

Age ___________________ Gender _______________________

Major __________________ Year in School _______________________

Please place a check mark next to the classes that you completed in **high school**:

____Biology ____Chemistry ____Physics ____Psychology ____Anatomy

____Physiology

Other Science related classes (please list):_________________________________________

Please place a check mark next to the classes that you completed or are taking in **college**:

____Biology ____Chemistry ____Physics ____Psychology ____Anatomy

____Physiology

Other Science related classes (please list):_________________________________________

Please place a check mark next to the statements that apply to you:
______ I have participated in science programs or research fairs.
______ Science is my favorite subject in school.
If so, please list favorite topic (e.g., biology, chemistry, etc.): __________________
______ I earned mostly As and Bs in my science classes in high school and college.
______ I enjoy watching science documentaries about biology or anatomy.
______ I would like to have a career in a science-related field.
______ I can name most of the components of the circulatory system from memory.
______ I have attended a course on cardiopulmonary resuscitation (CPR) training.
______ I sometimes find myself on the internet looking up science related topics.
______ I can draw the structure of the eukaryotic cell from memory.
______ I took advanced (AP, IB, Honors) science classes in high school.

Please rate your knowledge of the human body:
______ Very high
______ Somewhat high
______ Average
______ Somewhat low
______ Very low

# Appendix B

## Experiment 2: Pre-lesson Questionnaire

Age ____________________ Gender ______________________ Race_______________

Major _____________________ Year in School _______________________

Please place a check mark next to the classes that you completed in **high school**:

____World History ____U.S. History. ____European History ____Western Civilizations

____ Asian History

Other History related classes (please list):___________________________________________

Please place a check mark next to the classes that you completed or are taking in **college**:

____World History ____U.S. History. ____European History ____Western Civilizations

____ Asian History

Other History related classes (please list): __________________________________________

Please place a check mark next to the statements that apply to you:
______ I took advanced (AP, IB, Honors) history classes in high school.
______ History is my favorite subject in school.
______ I earned mostly As and Bs in my history classes in high school and college.
______ I enjoy watching shows on The History Channel.
______ I enjoy watching non-fictional war movies/documentaries.
______ I enjoy watching fictional war movies.
______ I would like to have a career in a history-related field.
______ I can describe several battles that occurred during World War II.
______ I can describe the Asia-Pacific War.
______ I sometimes find myself on the internet looking up history related topics in my free time.
______ I can draw a map of the major territories inhabited by parties involved during WWII.

Please rate your knowledge of World War II:
______ Very high
______ Somewhat high
______ Average
______ Somewhat low
______ Very low

## Appendix C

## Post-lesson Questionnaire for Experiment 1 and 2

Please rate each of the following statements about the lesson you just viewed.

| | Strongly disagree | | | | | Strongly agree | |
|---|---|---|---|---|---|---|---|
| 1. The lesson was difficult. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2. I feel like I have a good understanding of the material. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 3. I worked hard during the lesson. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4. The lesson was mentally demanding. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5. I could teach someone else what I just learned. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 6. I put a lot of effort in the lesson. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 7. I enjoyed learning this way. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8. I would like to see more of this kind of lesson in my classes. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 9. I was interested in this topic before the experiment. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 10. This lesson made me want to learn more about this topic. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 11. I felt motivated to understand the lesson. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 12. This lesson was useful to me. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 13. It was hard to pay attention during the lesson. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 14. I tried to remember the information in the order presented. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 15. I felt distracted during the lesson. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 16. I was trying to make sense of the material. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Please rate each of the following statements about the lesson you just viewed.

| | Strongly disagree | | | | | | Strongly agree |
|---|---|---|---|---|---|---|---|
| 17. I was working to memorize the information. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 18. I was trying to make connections between the material and things I already know. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 19. I was working on understanding the lesson. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 20. My mind was not on the lesson. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 21. I was trying to learn the main facts from the lesson. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 22. I felt that I was able to control events in the environment. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 23. The environment was responsive to actions that I performed. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 24. My interaction with the environment seemed very natural. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 25. The visual aspect of the virtual environment was compelling. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 26. I felt like the objects in the environment were really present and moving around me. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 27. I felt like I was really in the environment. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 28. I felt like I was immersed (or included in) and interacting with the computer environment. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Please rate how strongly you felt the following emotions overall during the lesson:

| | Not at all | | | | | Very Strongly | |
|---|---|---|---|---|---|---|---|
| 1. I felt happy | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2. I felt excited | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 3. I felt calm | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4. I felt content | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5. I felt sad | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 6. I felt angry | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 7. I felt frustrated | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8. I felt bored | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Do you have any additional comments about your experience with the lesson?

# Appendix D

## Experiment 1: Post-test with Correct Answers

1. Which molecule(s) cannot pass through a cell’s membrane without the help of a receptor protein?
    a. Water
    b. Oxygen
    **c. Glucose**
    d. All of the above

2. List the order of filaments in the cytoskeleton from narrowest to widest.
    **a. Microfilament, intermediate filament, microtubule**
    b. Intermediate filament, microtubule, microfilament
    c. Microtubule, intermediate filament, microfilament
    d. Microfilament, microtubule, intermediate filament

3. Which type of cells in the bloodstream are more commonly known as platelets?
    a. Leukocytes
    b. Monocytes
    c. Erythrocytes
    **d. Thrombocytes**

4. In what process are ribosomes involved?
    a. ATP production
    **b. Protein synthesis**
    c. Protein transportation
    d. Both B and C

5. Up to how many steps can kinesin motor protein take in a second?
    a. 50
    **b. 100**
    c. 150
    d. 200

6. Along which filament of the cytoskeleton does the kinesin motor protein move?
    a. Microfilament
    b. Intermediate filament
    **c. Microtubule**
    d. All of the above

7. In which direction do arteries carry blood?
    a. Toward the heart
    **b. Away from the heart**
    c. Both of the above
    d. Depends on the location of the artery

8. In what structure do proteins get transported?
    - **a. Vesicle**
    - b. Rough endoplasmic reticulum
    - c. Kinesin motor protein
    - d. Ribosome

9. Which type of blood cell can turn into a macrophage?
    - a. Erythrocyte
    - **b. Monocyte**
    - c. Thrombocyte
    - d. None of the above

10. What is converted into water and carbon dioxide in the mitochondria during ATP production?
    - a. Actin
    - b. Kinesin
    - c. ADP
    - **d. Pyruvate**

11. Where are mitochondria located?
    - **a. Free floating in the cytoplasm**
    - b. On the rough endoplasmic reticulum
    - c. In the nucleus
    - d. Attached to vesicles

12. Which type of blood cell takes up almost of half of the blood’s volume?
    - a. Leukocytes
    - b. Monocytes
    - **c. Erythrocytes**
    - d. Thrombocytes

13. What is the main source of energy in the cell?
    - a. DNA
    - b. ADP
    - **c. ATP**
    - d. Both B and C

14. What transports larger structures within the cell?
    - a. Ribosomes
    - b. Vesicles
    - **c. Kinesin motor protein**
    - d. Cytoplasm

15. What makes up the first line of our immune system?
    a. White blood cells
    **b. Antibodies**
    c. Leukocytes
    d. All of the above

16. Which structure allows entry and exit of larger molecules into the nucleus?
    a. Cytoskeleton
    b. Protein filaments
    c. Sodium potassium pump
    **d. Ribosomes**

17. What could cause kinesin motor protein to move more slowly?
    - **Larger molecules being carried**
    - **Lack of ATP**
    - **Something wrong with the microtubule that it walks on**

18. List some possible outcomes after a sudden decrease of red blood cells in the blood stream.
    - **Blood would look less red**
    - **Less transportation of oxygen to the rest of the body**
    - **If due to a tear, then monocytes would exit the blood vessel to search for foreign invaders and platelets would repair the tear**
    - **Fatigue/feeling tired or anemia**

19. What would happen to proteins if the rough endoplasmic reticulum was not functioning properly?
    - **Proteins may not be synthesized from the ribosomes on the RER**
    - **Defective proteins may be synthesized**
    - **Vesicles may not be made from the RER**
    - **Proteins may not be transported by vesicles**

20. Describe what happens on a cellular level after you get a cut on your finger.
    - **Blood vessel/tissue/skin tears**
    - **White blood cells (leukocytes) or monocytes exit the blood vessel to search for foreign invaders**
    - **Platelets (thrombocytes) stop bleeding at the damaged blood vessel**
    - **White blood cells/antibodies prevent viruses from coming into cells**

## Appendix E

## Experiment 2: Post-test with Correct Answers

1. On which Australian city did Japan launch a surprise attack on during WWII?
   a. Sydney
   **b. Darwin**
   c. Brisbane
   d. Perth

2. Who did Australia first send to New Guinea?
   **a. Militiamen**
   b. Trained soldiers
   c. The Australian Navy
   d. All of the above

3. Where did the Australians first arrive in New Guinea?
   **a. Port Morseby**
   b. Kokoda
   c. Buna
   d. Owen Stanley Range

4. Where did the Japanese first arrive in New Guinea?
   a. Port Morseby
   b. Kokoda
   **c. Buna**
   d. Owen Stanley Range

5. What were the colors of the Australian 39th Battalion?
   a. Blue and orange
   **b. Brown and red**
   c. Green and purple
   d. Black and white

6. What happened on July 28, 1942?
   a. The Japanese defeated the Australians in the Own Stanley Range.
   b. The Australians sent reinforcements to Port Morseby.
   c. The Japanese arrived at Buna.
   **d. The Australians took control of Kokoda.**

7. Who joined the Australian forces at Kokoda?
   a. The Papuan Infantry Battalion
   b. The Australia-New Guinea Administrative Unit
   c. Volunteer men from Papua
   **d. All of the above**

8. What were the Australian New Guinea Administrative Unit responsible for?
   **a. Organizing labor and supplies**
   b. Navigating the rough terrain
   c. Housing the Australian troops
   d. All of the above

9. Why did the Australians retreat from Kokoda?
   **a. Their leader, Lieutenant Colonel Owen, was killed.**
   b. They were outnumbered by the Japanese.
   c. They ran out of food.
   d. Both B and C

10. Who was the Australian leader at Mission Ridge?
   a. General Douglas MacArthur
   b. Lieutenant Colonel Owen
   **c. Brigadier Arnold Potts**
   d. General Thomas Blamey

11. What happened to the 2/27th AIF Battalion at Brigade Hill?
   a. They were defeated by the Japanese.
   **b. They were cut off by the Japanese and had to retreat.**
   c. They took Brigade Hill without fighting.
   d. None of the above

12. How did the Australian soldiers feel about the decision to withdraw from Mission Ridge and Brigade Hill?
   a. Relieved because they were hungry and tired
   **b. Upset because had been fighting hard for Australia**
   c. Angry because they didn't like the Australian leaders
   d. None of the above

13. What happened when the Japanese tried to send reinforcements to Buna?
   **a. They were defeated by air attacks**
   b. They were defeated by land attacks
   c. They were successful and were able to send food and supplies to their men
   d. Both A and B

14. Why did the Japanese leaders decide to put a hold on capturing Port Morseby?
   a. They ran out of supplies
   b. They had already captured Kokoda
   **c. They suffered a great loss somewhere else**
   d. The settled a treaty with Australia

15. How many days did it take the Australians to travel from Ioribaiwa to Kokoda?
    a. 12 days
    b. 24 days
    **c. 36 days**
    d. 48 days

16. What happened on November 3, 1942?
    a. The Japanese surrendered to the Australians at Kokoda
    b. The Japanese held their ground at Kokoda
    c. The Australians retreated from Kokoda
    **d. An Australian General raised a flag at Kokoda**

17. Why was Kokoda an important area to capture?
    - **Only place an aircraft could land for 100km**
    - **Could be used to deliver reinforcements and supplies**
    - **The plateau itself is a strong defensive position**

18. How did the terrain of the Owen Stanley Range impact both the Japanese and Australian troops?
    - **The terrain was rough and included narrow tracks, rivers, forest, dense jungle, and mountains**
    - **The terrain was difficult to traverse/slowed down troops on both sides**
    - **It was hard to navigate/easy to get lost**
    - **There were little resources for food and shelter**

19. How did the Australians gain control of Kokoda the second time?
    - **The Japanese retreated because of a major defeat**
    - **The Japanese were defeated at a different location**
    - **The Japanese were unable to get resources reinforcements**
    - **There was no fighting at Kokoda/ it was peaceful**

20. Describe how this portion of WWII, as well as the outcome of WWII, could have been impacted if the Japanese gained control over Kokoda.
    - **The Japanese would have had a more strategic location at Papua New Guinea because of the airstrip; could get more supplies and reinforcements**
    - **The Japanese may have been able to control more/all of Papua New Guinea, including Port Moresby**
    - **It may have extended fighting between the Australians and Japanese**
    - **The Australians may have been weakened if they had lost at Kokoda; could have lost New Guinea/Australia**
    - **The Japanese may have succeeded in their goal to control Australia**
    - **The Japanese may have had more power in WWII overall/could have defeated other countries**

# Appendix F

## Experiment 1: Practice Questions and Answers

1. Which blood cell carries oxygen?
   **Red blood cell or erythrocyte**

2. Name a molecule that can pass through a cell membrane freely.
   **Water or Oxygen**

3. What are the three types of strands that make up the cytoskeleton of the cell?
   **Microfilaments, intermediate filaments, and microtubules**

4. Which process turns DNA into RNA?
   **Transcription**

5. Where are ribosomes located?
   **On the RER**

6. How can viruses enter a cell?
   **By getting past white blood cells/antibodies and using a "counterfeit key"**

## Appendix G

## Experiment 2: Practice Questions and Answers

1. What event at Darwin led to Australia sending troops to New Guinea?
   **It was bombed or attacked by the Japanese**

2. Who took control of Kokoda first?
   **Australians**

3. What was the key strategic feature of Kokoda?
   **The airstrip or the plateau itself, some kind of description about it being a defensive position**

4. What did Brigadier Arnold Potts lead the Australians to do at Brigade Hill?
   **They retreated**

5. What happened to the Japanese troops at Guadalcanal?
   **They suffered a serious defeat**

6. Where did Japanese convoy retreat to?
   **Oivi or New Britain**