Extending the Cognitive-Affective Theory of Learning with Media in Virtual Reality Learning: A Structural Equation Modeling Approach

Article *in* Journal of Educational Computing Research · November 2021

DOI: 10.1177/07356331211053630

4 authors, including:

Wen Huang
Southern Methodist University

Rod Roscoe
Arizona State University

Scotty D Craig
Arizona State University

Original Research Article

Journal of Educational Computing Research
2021, Vol. 0(0) 1–36

DOI: 10.1177/07356331211053630
journals.sagepub.com/home/jec
SAGE

Extending the Cognitive-Affective Theory of Learning with Media in Virtual Reality Learning: A Structural Equation Modeling Approach

Wen Huang[1,2], Rod D. Roscoe[3], Scotty D. Craig[3], and Mina C. Johnson-Glenberg[4]

Abstract

Virtual reality (VR) has a high potential to facilitate education. However, the design of many VR learning applications was criticized for lacking the guidance of explicit and appropriate learning theories. To advance the use of VR in effective instruction, this study proposed a model that extended the cognitive-affective theory of learning with media (CATLM) into a VR learning context and evaluated this model using a structural equation modeling (SEM) approach. Undergraduate students (n = 77) learned about the solar system in a VR environment over three sessions. Overall, the results supported the core principles and assumptions of CATLM in a VR context (CATLM-VR). In addition, the CATLM-VR model illustrated how immersive VR may impact learning. Specifically, immersion had an overall positive impact on user experience and

[1]Department of Teaching and Learning, Simmons School of Education and Human Development, Southern Methodist University, Dallas, TX, USA
[2]Engineering Education Systems and Design, The Polytechnic School, Ira A. Fulton Schools of Engineering, Arizona State University, Mesa, AZ, USA
[3]Human Systems Engineering, The Polytechnic School, Ira A. Fulton Schools of Engineering, Arizona State University, Mesa, AZ, USA
[4]Department of Psychology, Arizona State University, Tempe, AZ, USA

Corresponding Author:
Wen Huang, Department of Teaching and Learning, Simmons School of Education and Human Development, Southern Methodist University, 6425 Boaz Lane, Dallas, TX 75205, USA.
Email: wenhuang@smu.edu

motivation. However, the impact of immersion on cognitive load was uncertain, and that uncertainty made the final learning outcomes less predictable. Enhancing students' motivation and cognitive engagement may more directly increase learning achievement than increasing the level of immersion and may be more universally applicable in VR instruction.

Keywords

virtual reality, CATLM, structural equation modeling, theory extending, multimodal learning

Introduction

Virtual reality (VR) is a fully-immersive 3-D multimedia environment where individuals can interact with a computer generated world (Aukstakalnis & Blatner, 1992; Milgram & Kishino, 1994; Onyesolu & Eze, 2011; Oxford, 2019). VR offers powerful affordances related to 3-D immersion, spatial representations, and multi-sensory cues (Salzman et al., 1999; Shin, 2017), and these characteristics allow learners to experience real or imagined environments that might be otherwise inaccessible (Huang & Roscoe, 2021). Prior studies have demonstrated learning from VR when studying microscopic processes (e.g., chemical reactions; Bennie et al., 2019), large-scale processes (e.g., solar system events; Huang et al., 2021), dangerous processes (e.g., emergency in mines; Grabowski, 2019), or processes too difficult or expensive to explore in real life (e.g., construction; Angulo & Velasco, 2014). In addition, other research has found that VR can enhance spatial knowledge, intrinsic motivation, and transfer of new abilities to real-world situations (Huang, 2019; Ragan et al., 2015; Timcenko et al., 2017).

However, in a recent review of 59 publications (published between 2009 and 2018) on VR learning applications for higher education, Radianti et al. (2020) observed that about two-thirds of these articles did not incorporate an explicit learning theory. This issue is crucial because design and implementation of effective educational technology interventions is enhanced by grounding such work in testable theoretical models of learning (Granić & Marangunić, 2019; Valverde-Berrocoso et al., 2020). These models provide insight into fundamental learning processes and outcomes, which in turn inform decisions about technology features and pedagogy. Radianti et al. (2020) further argued that many extant learning theories do not include elements specifically related to learning with VR (e.g., immersion, presence, or embodiment). Thus, a substantive amount of work in the field of VR learning may be either unguided by a clear model of learning or may be using models that are incomplete. To advance the use of VR in effective instruction, it may be useful to extend and test learning theories for use in that context.

The cognitive-affective theory of learning with media (CATLM) is a popular model among VR learning researchers and developers, and represents a very plausible and straightforward candidate for elaboration based on VR learning principles (Mayer, 2020; Moreno & Mayer, 2007). In brief, CATLM is a theory for understanding learning with instructional materials beyond words and images (e.g., manipulatives). This theory articulates mechanisms for meaningful learning that may occur when learners directly interact with the instructional system (e.g., dialogue, control, and manipulation) in a multimodal learning experience (Moreno & Mayer, 2007). The advantages of including different types of sensory interactions (e.g., auditory, visual, and tactile) and different domains of human functioning (e.g., cognition, motivation, and emotion) make CATLM seem an ideal foundation for understanding the VR learning process. CATLM has been credited in the design of computer simulations, computer-based instruction, online learning, and serious games (e.g., Domagk et al., 2010; Krämer & Bente, 2010; Plass & Schwartz, 2014; Ritterfeld et al., 2009; Sosa et al., 2011). With regards to VR learning, CATLM has been used to define the parameters and evaluation of user experience and user-centered instructional design, such as visualizations (Birt et al., 2015), human-robot collaboration (Shu et al., 2019), teaching concepts, and procedures (Ta, 2018); and to inform pedagogy for increasing learning performance, such as knowledge construction (Spek et al., 2008), knowledge transfer (Petersen et al., 2020), and behavior involvement (Sajjadi et al., 2018).

Importantly, as a theory developed before the most recent boom of VR technologies, CATLM may not fully encompass VR learning for two reasons. First, CATLM assumes that motivation affects learning in an interactive multimodal environment (e.g., VR) (Leutner, 2014). However, CATLM does not include several essential constructs of VR environments (e.g., immersion, presence, and embodiment) that may have connections with motivation in VR learning (Johnson-Glenberg, 2018; Parong & Mayer, 2018). As a result, the model does not specifically articulate the claim that motivation would be important for VR learning. Second, in the field of VR, the model has been used to influence design principles but has not been fully validated (Makransky et al., 2019; Mayer, 2020). There is a lack of evidence assessing how and whether elements of VR design (e.g., level of immersion) and elements of CATLM (e.g., cognitive load) indeed interact to influence learning.

Current Study

The purposes of this study are to (a) extend CATLM for VR learning and (b) to evaluate the applicability and validity of the resulting model. Specifically, we briefly summarize CATLM principles and assumptions and then suggest extensions to this model based on VR learning environments. This review results in an updated theoretical model with multiple testable hypotheses. We then describe a study in which undergraduate participants learned with different VR formats over several sessions. Data were collected on participants' learning performance along with their subjective experiences of presence, embodiment, motivation, and cognitive demands. Structural equation modeling (SEM) methods are then used to explore relations among variables and test whether these connections align with our extended model of CATLM for VR.

Extending Cognitive-Affective Theory of Learning with Media for Virtual Reality Contexts

Overview

A total of nine variables and 20 hypothesized relationships are articulated within our extended model. Among these concepts, five are sourced directly from CATLM (i.e., prior knowledge, cognitive load, cognitive engagement, motivation, and learning achievement) and four are added from the literature on VR learning (i.e., VR format, embodiment, presence, and time). In this research model, three concepts (i.e., VR format, time, and prior knowledge) are defined as external variables. The remaining concepts are recognized as constructs. It is worth noting that cognitive load is operationalized here as a two-level construct with two subordinate constructs: extraneous cognitive processing and essential cognitive processing. Figure 1 visually summarizes

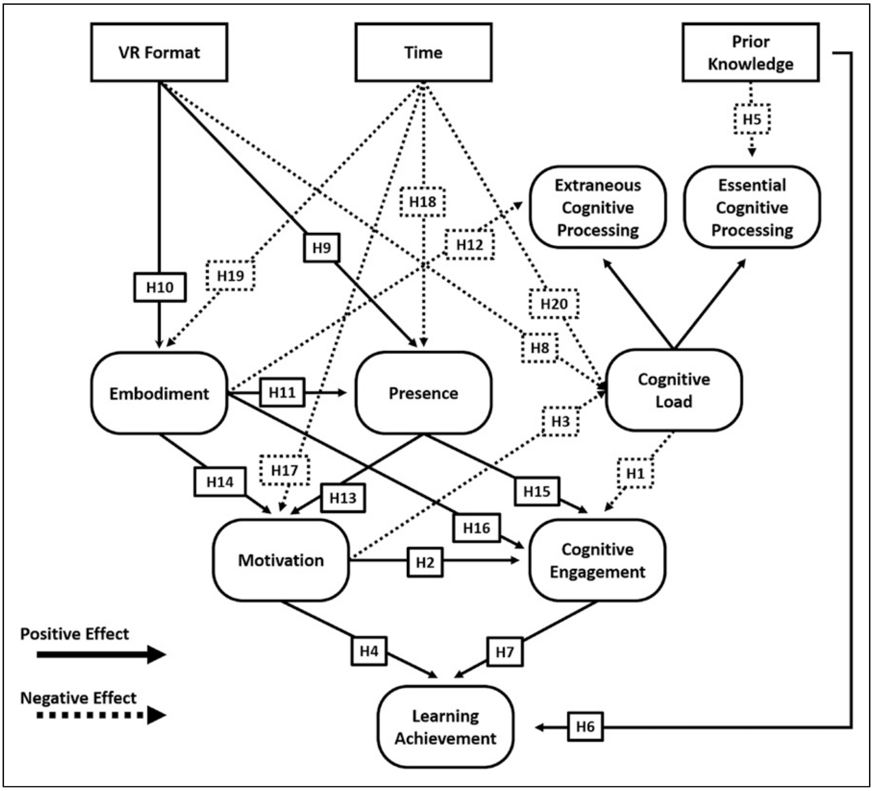

Figure 1. Hypothesized model.

the hypothesized research model. Subsequent sections provide more specific details and explanations of the model components.

Foundations of Cognitive-Affective Theory of Learning with Media

CATLM (Moreno & Mayer, 2007) provides a coherent perspective for learning in interactive multimodal environments—learning environments that may combine text, narration, illustrations, music, and/or sound effects. Prior articulations of CATLM are based upon evidence-based observations of human cognition and affect, which give rise to several core constructs that directly or indirectly influence learning: *cognitive load*, *cognitive engagement*, *motivation*, and *prior knowledge*.

Cognitive load and cognitive engagement. One assumption of CATLM is that humans possess somewhat independent channels for processing information of different modalities, such as visual and auditory stimuli (Baddeley, 1992; Moreno & Mayer, 2007). In addition, these channels are limited with respect to the amount and duration of information that can be processed at a given time (Moreno & Mayer, 2007; Sweller, 1999). Importantly, because cognitive capacities and resources (e.g., working memory) are limited, learning environments and tasks that exceed these limitations will result in cognitive overload that stymies productive processing (Mayer, 2017).

With regard to design, various formulations have described two or three aspects of cognitive processing that must be considered. Essential processing refers to processes inherent to the task and learning (e.g., noticing and selecting relevant information), and extraneous processing occurs when learners must contend with irrelevant information or distracting details that contribute minimally to the task (e.g., a confusing interface). Both essential and extraneous processing represent *cognitive load* due the design of the multimodal learning environment (Kalyuga, 2011). In contrast, generative processing entails *cognitive engagement* that extends beyond inherent requirements in order to facilitate recall and comprehension (e.g., self-explaining and self-questioning learning strategies) (Chi et al., 2018; Kalyuga, 2011; Mayer, 2020; Sweller, 2010). Deeper cognitive engagement may also include metacognitive processes for self-evaluation and regulation (Adadan, 2020; McGuinness, 1990; Moreno & Mayer, 2007). Such engagement represents learners' in-the-moment learning experience (Boyle et al., 2012; Shernoff & Strati, 2011), sometimes characterized by higher intensity of arousal, attention, involvement, and persistence (Cordova & Lepper, 1996; Garris et al., 2002; Parker & Lepper, 1992).

H1: Cognitive load (i.e., essential and extraneous processing) has a direct and negative effect on cognitive engagement (i.e., generative processing). When learners experience higher cognitive load, it may impede their ability to think deeply and constructively about the material.

Motivation and prior knowledge. Another assumption is that meaningful learning requires learners to invest conscious effort in cognitive processes ranging from basic attention to higher-level generation and reasoning (Lee & Heeter, 2017; Mayer & Moreno, 2003). Long-term memory consists of dynamically organized knowledge

(Sweller et al., 1998; Tulving, 1977) that results from and is transformed by the aforementioned reasoning processes. Such cognitive engagement is influenced by several factors, including learners' motivation and goals (Mayer, 2014; Pintrich, 2003), and learners' prior knowledge, abilities, and skills within specific media (Kalyuga et al., 2003; Mashfufah et al., 2019).

According to CATLM, *motivation* positively influences learners' cognitive engagement (i.e., generative processing) (Kuldas et al., 2014; Roets et al., 2008; Schnotz & Kürschner, 2007). In academic settings, motivation can be defined as a desire to learn or improve achievement (Mayer, 2020), and strongly motivated individuals are more likely to use strategies (e.g., Holmes, 2018; Schiefele, 1991) to reduce extraneous processing, manage essential processing, or foster generative processing (Mayer, 2020). Motivation can reduce the perceived effort (i.e., perceived cognitive load) associated with a learning task (Song et al., 2019), and a higher degree of cognitive engagement is sometimes associated with a flow state (Csikszentmihalyi, 1997; Whitton, 2011). Motivation also seems to maintain attention to the learning task and encourage persist when difficulties are encountered (Svinicki, 2004). There is an overall positive relationship between motivation and learning performance in classroom teaching (Keller, 2010; Lee et al., 2010; Lepper & Iyengar, 2005; Pintrich, 2003).

H2: Motivation has a direct and positive effect on cognitive engagement. Highly motivated learners are more likely to choose and implement generative learning strategies.

H3: Motivation has a direct and negative effect on cognitive load. Learners who are more motivated may perceive less difficulty, perhaps because the sense of challenge is rewarding or because they are in a flow state.

H4: Motivation has a direct and positive effect on learning achievement. Learners with a high motivation are likely to have better performance in post-learning assessments.

Within CATLM, prior knowledge has at least two effects on learning in an interactive multimodal environment. First, prior knowledge supports learners in identifying and selecting useful information (Mashfufah et al., 2019). Second, prior knowledge facilitates constructing new knowledge based on new information (Altas, 2015). Learners with more extensive domain knowledge and expertise can better pay attention to new relevant information and build upon their existing knowledge (Barab et al., 2007). In other words, their required essential cognitive processing is reduced. In addition, prior VR learning research shows that pre-training of background information can promote further knowledge gains (Petersen et al., 2020). Overall, instructional designers can implement a variety of methods to help learners activate and apply their prior knowledge and pursue more optimal cognitive and metacognitive strategies in interactive multimodal learning environments (McGuire, 2015; Moreno & Mayer, 2007) such as guided activities (e.g., Plass & Kaplan, 2016), problem solving (e.g., Hou & Li, 2014), self-explaining (e.g., Hsu et al., 2012), prompts for reflection (e.g., McConnel, 2018), explanatory feedback (e.g., Johnson et al., 2017), pace control (e.g., Biard et al., 2018), and pre-training (e.g., Petersen et al., 2020; Pineda, 2015).

H5: Prior knowledge has a direct and negative effect on cognitive load. Specifically, when learners possess stronger prior knowledge, such knowledge can be leveraged to reduce essential processing demands.

H6: Prior knowledge has a direct and positive effect on learning achievement. Students learn more effectively when they possess broader or deeper prior knowledge with which to connect or interpret new information.

H7: Cognitive engagement (i.e., generative processing) has a direct and positive effect on learning achievement. Students learn better when they intentionally select and organize information to construct new knowledge.

Learning with Virtual Reality

A number of VR developers and scholars have drawn upon principles relevant to CATLM to guide their work (e.g., Makransky & Lilleholt, 2018; Schrader & Bastiaens, 2012; Tcha-tokey et al., 2018), and VR learning appears to be within the scope of CATLM (see Moreno & Mayer, 2007). Nonetheless, unique aspects of VR inform several constructs that might be added to further tailor CATLM to the VR context. Specifically, the immersive nature of VR can contribute to learners' sense of existing within the virtual environment (e.g., Shin, 2017; Shin et al., 2013; Tcha-tokey et al., 2018), and VR tools (e.g., motion tracking and headsets) can afford varying degrees of physical sensations, movements, gestures, and interaction with virtual objects (Johnson-Glenberg, 2018). Overall, considerations of VR learning suggest four elaborations of CATLM for VR: *VR format*, *presence*, *embodiment*, and *time*.

VR format. A variety of VR systems and corresponding applications have been released in the recent years. In general, these systems can be categorized into three levels of sophistication (see Mcmillan et al., 2017). The simplest format typically includes a headset that uses a smartphone to present the VR environment without additional features or functionality (e.g., Google Cardboard). More advanced formats use headset-and-smartphone arrangements or standalone headsets (i.e., head-mounted displays) along with additional functions, such as hand-held tools for navigation and interaction (e.g., Samsung Gear VR and Oculus Rift Go). Finally, the most sophisticated formats include high-end headsets that can track individuals' positions, multiple controls for interaction and manipulation (e.g., two hand-held controllers), and motion tracking (e.g., Oculus Rift and HTC Vive). As technologies become more affordable and accessible, the simplest format has become increasingly obsolete—the mid-range and high-end VR systems continue to be of interest.

Importantly, the technical differences between mid-range (e.g., Oculus Go) and high-end (e.g., Oculus Rift) VR systems result in differences in representational fidelity and flexible interactions between the learner and the virtual world. That is, high-end systems allow virtual environments to appear more realistically, with inputs across more modalities (e.g., sight, sounds, and touch), and can be experienced via more natural movements (e.g., grasping and rotating objects). Such features afford a deeper level of immersion, namely the capability of "delivering an [……] illusion of reality to

Table 1. A Comparison of the Oculus Go and the Oculus Rift VR Systems.

	Oculus Go	Oculus Rift
Technical Features		
Display	2560 × 1440, LCD	2160 × 1200, OLED
Refresh rate	60–72 Hz	90 Hz
Field of view	100°	110°
Body-tracking	3 degrees of freedom	6 degrees of freedom (room scale)
Controller	Unimanual controller	Bimanual controllers
Audio	Headphones purchased separately	Integrated over-ear headphones with 3-D directional audio support
Supported modalities		
Auditory	Yes	Yes
Visual	Yes	Yes
Tactile	No	Yes
Manipulation	No	Yes

the senses of a participant" (Slater & Wilbur, 1997, p. 606). Table 1 summarizes technical differences between the Oculus Go and Oculus Rift systems, which correspond to "moderate immersion" and "higher immersion" VR, respectively.

In addition, previous research showed that gestures and manipulation have the potential to help students learn new knowledge (Goldin-Meadow et al., 2001; Hu et al., 2015; Paas & Sweller, 2012). One explanation is that using two (or more) modalities to encode meaning magnifies the learner's available cognitive capacity and reduces cognitive demand in a single processing channel (Goldin-Meadow, 2011). Due to more flexible interactions, high-end VR systems may offer an advantage for reducing cognitive load compared to mid-range VR systems.

H8: VR format has a direct and negative effect on cognitive load (i.e., essential and extraneous processing). Higher levels of immersion and interaction reduce the effort needed to learn.

Presence and embodiment. Presence and embodiment are regarded as two essential affordances of VR in learning (Johnson-Glenberg, 2018). A *sense of presence* is defined as "a state of consciousness, the (psychological) sense of being in the virtual environment" (Slater & Wilbur, 1997, p. 607) and relates to a feeling of "being there" or place illusion (Slater, 2009). A *sense of embodiment* refers to the feeling of being inside of and controlling a virtual body within a VR environment (Kilteni et al., 2012), and requires some degree of congruence between predicted and actual sensory feedback (Johnson-Glenberg, 2018; Kokkinara et al., 2016; Sato & Yasuda, 2005).

H9: VR format has a direct and positive effect on presence. Higher levels of immersion lead to a stronger sense of "being within" the virtual world.

H10: VR format has a direct and positive effect on embodiment. Higher levels of immersion enable more control and movement that aligns with natural actions.

Feelings of presence and embodiment are related to one another, as well as other subjective experiences of effort and motivation. Individuals who engage in more bodily movements with an identified virtual body report a higher degree of presence in a VR environment (Slater et al., 1995, 1998). Embodiment is related to the experience of autonomy and agency when controlling the virtual body (Johnson-Glenberg, 2018; Kilteni et al., 2012), and a high degree of embodiment potentially offers learners more choices for learning the subject matter. Due to this increased learning flexibility, learners' perceived mental effort may be decreased. Finally, feelings of presence and embodiment in VR environments are novel for many learners. For example, learners can see the structure of the solar system and freely manipulate a planet in a VR environment, both of which are impossible in the real world. Due to curiosity, learners may be intrinsically motivated to learn (Malone & Lepper, 1987; Wade & Kidd, 2019).

H11: Embodiment has a direct and positive effect on presence. The ability to move and interact with a virtual world contributes to the feeling of "being within" that world.

H12: Embodiment has a direct and negative effect on extraneous cognitive processing. The ability to interact with the virtual world in a more natural manner reduces perceived effort.

H13: Presence has a direct and positive effect on motivation. The experience of "being within" a novel, unfamiliar, or impossible world may inspire intrinsic interest and curiosity.

H14: Embodiment has a direct and positive effect on motivation. The freedom to move and interact with a virtual world more naturally may inspire intrinsic interest and curiosity.

Furthermore, both sense of presence and sense of embodiment have the potential to impact cognitive engagement and learning. The degree of cognitive engagement in a virtual environment is partially determined by the user's evaluation of the balance between personal skills and any challenges encountered during the learning process (Takatalo, 2002). Thus, as a concept related to individuals' perception of their relationship between self and the virtual world (Kilteni et al., 2012), presence can be recognized as a prerequisite for cognitive engagement. The possible effect of embodiment on cognitive engagement in a VR environment is attributed to the user interface. A more "natural" user interface (i.e., familiar behaviors with controls that align to everyday movements and interactions) enhances individuals' sense of embodiment and facilitates behaving in a virtual environment similar to the real world (Bianchi-Berthouze, 2013; Brondi et al., 2015; Lindgren et al., 2016). When students learn in more-embodied settings, they report a higher degree of cognitive engagement than less-embodied settings (Lindgren et al., 2016).

H15: Presence has a direct and positive effect on cognitive engagement (i.e., generative processing). A sense of "being within" the virtual world may enable further exploration, analysis, and questioning about ideas encountered in that world.

H16: Embodiment has a direct and positive effect on cognitive engagement. The ability to interact with the virtual world in more natural ways (e.g., gestures) enables additional modalities for investigation, analysis, and inference.

Time. Learners' motivation is partially determined by their perception of novelty. If stimuli are unexpected and surprising, learners' motivation may increase immediately but then attenuate gradually over time due to the familiarity. This phenomenon is called the novelty effect (Koch et al., 2018).

Furthermore, motivation may influence learners' behaviors by moderating attentional resources (Engelmann & Pessoa, 2007; Robinson et al., 2012; Wicken, 1992). Sufficient, involuntary focused attention is the pre-condition of generating spatial presence in a VR environment (Bystrom et al., 1999; Wirth et al., 2007; Witmer & Singer, 1998). Similarly, motivation positively impacts embodied senses in an environment with few distractions (Zestcott, 2017). Thus, learners' sense of presence and embodiment may fluctuate as their motivation changes over time. Importantly, learners should also become more comfortable and adept at operating (and learning within) the VR system over time. Thus, the perceived difficulty or complexity of learning in a VR environment should decrease.

H17: Time has a direct and negative effect on motivation. Over time, initial novelty of VR may decline, leading to greater familiarity or boredom.

H18: Time has a direct and negative effect on presence. Over time, initial novelty may decline, and external distractions or concerns may become more salient. Users may also begin to attend to lack of fidelity of the virtual world (e.g., sight and sound) or lack of certain senses (e.g., smell and taste).

H19: Time has a direct and negative effect on embodiment. Over time, initial novelty may decline, and users may begin to attend to the incongruence between virtual actions and real-world actions. In addition, the experience of holding controllers and wearing a head-mounted display may become more salient.

H20: Time has a direct and negative effect on cognitive load. As users become more familiar with the VR controls and world, cognitive demands and effort to navigate that world should decrease.

Evaluating the Extended Model

A strong theoretical model should be subject to testing and validation (Jaccard & Jacoby, 2010; Littlejohn et al., 2017). Theory testing assesses whether a theory aligns with reality—whether concepts and their relationships specified in the theory can be observed in practical instances (Hogan & Schmidt, 2002; Littlejohn et al., 2017). This assessment is a key step for establishing validity (Littlejohn et al., 2017), and such testing usually involves not only inspecting connections among a set of concepts but also delving into the underlying processes (Colquitt & Zapata-Phelan, 2007). Previous researchers have used a variety of methods to test and extend different theories, these methods including simple *t*-tests (e.g., Brancheau, 1989; March & Woodside, 2005), regression analyses (e.g., Brancheau, 1989; Ward et al., 2014), correlation analyses (e.g., Brown, 2004), SEM (e.g., Hogan & Schmidt, 2002; Straub et al., 1995; Viswesvaran & Ones, 1995), meta-analyses, and case studies (e.g., Chukwudi et al., 2020; Løkke & Sørensen, 2014).

SEM is a multivariate technique originating from factor analysis and path analysis (Wang & Wang, 2012) that enables researchers to assess the quality of measurement and examine the relationships among constructs at the same time. Distinct from other statistical methods (e.g., regression, ANOVA, and correlation analysis), SEM takes into account measurement errors in the observed variables involved in a model. Thus, the parameter estimates may be more accurate and less biased. SEM is most appropriate when the researcher has multiple constructs with several indicators and some of these constructs simultaneously act as independent variables in one relationship but dependent variables in other relationships (Hair et al., 2009). SEM procedures are guided more by theory than by empirical results, and are considered a confirmatory analysis useful for testing and potentially confirming theory (Hair et al., 2009). For these reasons, the current work has selected SEM methods to initially assess the proposed extended model of CATLM for VR.

The model evaluation work was embedded in a larger research project implemented at a large public university in the southwest of the United States. Data were collected by survey questionnaires and knowledge quizzes in a three-session VR learning study in the fall semester of 2019, and intervals between sessions spanned 4–7 days.

Participants

A total of 77 undergraduate adult students with no history of serious cybersickness or motion sickness were recruited for the study. Among these recruited students, 60 did not have any VR experience before. Seven had zero to 5 h of experience, five had 5–10 h of experience, and the remaining five used VR weekly. According to the design of the larger research project, participants were randomly assigned to three groups. The first group (n = 23) completed their VR learning in the moderate immersion format across three sessions. The second group (n = 23) was identical to first group in the first two sessions. However, they switched to the higher-immersion format in the third session. The third group (n = 31) completed their VR learning in the higher-immersion format across three sessions. Due to attrition, only 50 participants completed all three sessions. Specifically, there were 16, 16, and 18 participants remained in these three groups, respectively. Total attrition was comparable among groups (30.4, 30.4, and 41.9%). Table 2 summarizes participant in this study. See Huang (2020) for demographic details in each group and session. It is worth noting that only remaining participants were included in the current analyses.

Materials

Learning content. The VR application *Titans of Space* (DrashVR, 2019) was used in this study. This application allows a player to engage in a virtual tour through the solar system, in which the player can learn the basic information (i.e., size, climate, and

Table 2. Demographic Information for Study Participants.

		Initial participants	Dropout	Attrition rate (%)	Remaining participants
Race	American Indian or Alaskan native	5	4	80.0	1
	Asian or Asian American	12	4	33.3	8
	Black or African American	6	0	0.0	6
	Hispanic or Latinx	16	4	25.0	12
	White or Caucasian	36	15	41.7	21
	Not reported	2	1	50.0	1
Gender	Male	52	18	34.6	34
	Female	25	9	36.0	16
Major	Engineering	45	12	26.7	33
	Science	14	6	42.9	8
	Business or humanities	18	9	50.0	9

components) of celestial bodies through reading, 3-D observation, and manipulation of objects during the virtual tour. *Titans of Space* has versions for both the Oculus Rift and the Oculus Go systems with the same learning content. Both systems share the same text resources and observable 3-D planet models. However, the Rift version differs from the Go version via its affordances for grabbing and rotating virtual subjects. Thus, learning with the Rift system represents a higher level of immersion (see Table 1). In this study, these two types of VR environments represented a moderate immersion format and a higher immersion format, respectively. More information about *Titans of Space* can be found on http://titansofspacevr.com/titansofspace.html.

To facilitate the implementation of the study, the content of *Titans of Space* was evenly separated to four "units," including Unit 1: Earth, Mercury, Venus, Mars and their moons; Unit2: Ceres, Jupiter and its moons; Unit 3: Saturn and its moons; and Unit 4: Chariklo, Uranus, Neptune, Pluto, and their moons. During the study, content in Unit 1 was used for training before the formal learning task in Session 1, Unit 2 was the learning task in Session 1, Unit 3 was the task in Session 2, and Unit 4 was the task in Session 3.

Measures. ***Prior knowledge.*** A pretest was used to assess participants' prior knowledge of the solar system. The test involved (a) ranking the eight planets from smallest to largest and (b) ranking these planets from closest to furthest from the Sun. This test was scored by assigning one point for each correct ranking, with a total possible score of 16. The test had a high test-retest reliability (see Huang et al., 2021).

Self-report measures. A questionnaire assessed presence, sense of embodiment, motivation, cognitive engagement, and cognitive load in each session. Items in the questionnaire were either directly or revised from validated instruments in previous literature (see Appendix A). The subscales of the current questionnaire showed acceptable internal reliability using Cronbach's α: Presence (α = 0.77), embodiment (α = 0.71), motivation (α = 0.91), and cognitive engagement (α = 0.79). A 0-100 sliding scale interface was used for all ratings, with "0" representing the lowest possible value (e.g., strongly disagree) and "100" representing the highest possible value (e.g., strongly agree). This type of slide is recommended for observing the within-person changes over time for its sensitivity (Grimm et al., 2016). The content validity was established by a panel of experts in the VR education and educational technologies fields. Face validity was established based on feedback from a panel of undergraduate and graduate students on the clarity of items.

Knowledge quizzes. Sets of knowledge quizzes were designed to assess learning achievement in each session. Specifically, each quiz comprised one recognition question, two recall questions, two understanding questions, and one evaluation question.

The *recognition question* required a participant to identify four celestial bodies they had learned in the unit using a drop-down menu of possible names. For scoring, one point was assigned per correct match. The *recall questions* asked the participants to describe "characteristics about the geology, climate, or orbit" about a given celestial body. For scoring, one point was assigned per correct fact. The stem of the *understanding question* was "What the life would be like on [a celestial body]." Responses were scored by assigning one point per valid argument. The argument had to be supported by factual details obtained in the VR materials, and there had to be a clear causal relationship between the claim and the fact(s). For example, "Enceladus would probably sustain life pretty well, with all the water one needs to live readily available, which could be converted to fuel machinery and breathe." The *evaluation question* asked the participants to answer "which celestial body do you think has greater potential for humans to live on" on two given objects (e.g., Mars and Jupiter). Scoring assigned one point per reasonable comparison or contrast. For example, "Humans have greater potential to live on Mars because it obtains more sunlight than Jupiter."

Two researchers independently scored participants' responses. The Pearson correlation coefficient for inter-rater agreement reached 0.90, $p < .001$ after they had scored 50% items. One rater then completed the remaining scoring.

Procedure

At the beginning of the first session, participants provided informed consent all IRB protocols were followed. All participants were required to complete the pretest and then participate a short training (6–8 min) with content Unit 1. After completing the training, participants began the Unit 2 learning task. In the following two sessions, they studied Unit 3 and Unit 4, respectively. Each formal learning unit (i.e., Unit 2, Unit 3, and Unit 4) required approximately 15–20 min. After the VR learning task in each session, participants completed a questionnaire and a knowledge quiz.

Data Analytical Approach

The data analyzed *via* SEM included the set of 50 participants who completed the entire study (i.e., all measures and sessions). Importantly, all participants were assessed three times during the study, thus resulting in a total of 150 data points. This data set satisfied the requirement of minimum size for SEM, which is about 100–150 (Anderson & Gerbing, 1988; Ding et al., 1995; Tinsley & Tinsley, 1987).

Analyses were conducted using Mplus 7.4 software. The SEM approach included a two-step analysis (Anderson & Gerbing, 1988). The first step was an evaluation of the measurement model. The second step was an estimate of the structural model. Four common goodness-of-fit indices were used to assess how well the model represented the data for both the measurement model and the structural model: the relative χ^2/df, root mean square error of approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and standardized root mean square residual (SRMR). The desired values were: $\chi^2/\text{df} < 3$ (Bagozzi & Yi, 1988), CFI and TLI >0.90 (Alwin & Hauser, 1975), and SRMR <0.08 (Hu & Bentler, 1999).

Additionally, the evaluation of the measurement model included establishing convergent validity and discriminant validity. Convergent validity requires that items belonging to the same factor converge or share a high proportion of variance in common (Hair et al., 2009). According to the recommendation of Hair and colleagues, convergent validity requires: (1) factor loadings ≥0.50, (2) average variance extracted (AVE) ≥ 0.50, and (3) composite reliability (CR) ≥ 0.70. Discriminant validity was established through a comparison between the AVE estimates for each factor and the squared inter-factor correlation associated with that factor.

Results

Descriptive Statistical Analysis

Before model evaluation, we screened all 150 data point and tested whether they met the criteria for conducting a multivariate analysis. This process included identifying outliers and testing for normal distribution. Two data point from the same participant were found to have more than two outliers on their variables, respectively. Because a data point having more than two outliers might seriously impact the overall analysis (Hair et al., 2009), the three data points from this participant were removed. The remaining 147 data points were included for the further analysis. Table 3 shows the descriptive statistics of the remaining data. The Mardia's multivariate kurtosis statistic was 27.77 ($p < 0.01$) and beyond the range of the recommended cut-off point (±3) suggested by Bentler (2006). This outcome suggested that the data were not multivariate normally distributed. Thus, the robust maximum likelihood method (MLR) was chosen as the estimation method for further analysis.

Measurement Model Evaluation

A confirmatory factor analysis (CFA) was conducted to assess the quality of the measurement model for the remaining 147 data points. The assessment results of the measurement model fitness were as follows: $\chi^2/\text{df} = 1.357$, RMSEA = 0.049, CFI =

Table 3. Descriptive Statistics of Data for SEM.

Factor	Item	Mean	Minimum	Maximum	*SD*
Presence	Pr_1	83.96	5	100	20.71
	Pr_2	70.66	0	100	27.04
	Pr_3	68.92	0	100	26.66
	Pr_4	60.21	0	100	29.31
Embodiment	Em_1	56.40	0	100	30.48
	Em_2	64.84	0	100	32.19
	Em_3	47.14	0	100	35.26
Motivation	Mo_1	90.01	31	100	13.99
	Mo_2	89.39	0	100	15.42
	Mo_3	91.28	30	100	12.24
Cognitive engagement	En_1	82.73	10	100	18.39
	En_2	83.86	25	100	16.78
	En_3	74.59	0	100	19.92
	En_4	81.08	16	100	19.39
	En_5	82.26	0	100	19.15
Essential cognitive processing	CI_1	32.46	0	100	25.37
	CI_2	37.38	0	100	27.24
Extraneous cognitive processing	CI_3	23.60	0	100	25.60
Learning achievement	Recognize	3.18	0	4	1.10
	Recall	4.26	0	11	2.30
	Understand	2.18	0	8	1.96
	Evaluate	1.07	0	3	0.93
Prior knowledge	Pretest	7.25	0	16	4.22

0.932, TLI = 0.918, SRMR = 0.069. All the values of these indices were in good model fit. Therefore, the goodness-of-fit statistics supported the measurement model.

For a single-item factor (i.e., extraneous processing), a fixed 1.00 loading and a fixed measurement error variance (i.e., zero) was specified for the factor's identification (see Hayduk & Littvay, 2012). Table 4 is an overview of the convergent validity for assessed factors. These data met or were very close to the guideline. Thus, convergent validity was confirmed.

In addition, all AVE estimates in Table 4 are greater than the corresponding squared correlation data in Table 5. In other words, the shared variance between factors were lower than the average variance extracted of the individual factors. This evidence supported the discriminant validity of the measurement model.

Structural Model Evaluation

Given the satisfactory results in the measurement model, we implemented SEM to test the causal relationship between factors in the research model. The research model reached a good fit (χ^2/df = 1.356, RMSEA = 0.049, CFI = 0.920, TLI = 0.906, SRMR =

Table 4. Summary of Measurement Scales.

Factor	Factor Loadings	AVE	CR
Presence	0.59–0.89	0.51	0.81
Embodiment	0.58–0.88	0.51	0.75
Motivation	0.82–0.93	0.78	0.91
Cognitive engagement	0.55–0.75	0.44	0.79
Cognitive load	0.68–1.00	0.73	0.84
Essential cognitive processing			
Extraneous cognitive processing			
Learning achievement	0.43–0.85	0.46	0.76

Table 5. Standardized Factor Correlation Matrix.

	1	2	3	4	5	6	7	8
2	0.45***	—						
3	0.42***	0.41***	—					
4	0.31**	0.21*	0.47***	—				
5	−0.08	0.04	−0.30*	−0.70***	—			
6	−0.09*	−0.06	−0.23*	−0.48***	—	—		
7	−0.16*	−0.12	−0.36*	−0.73***	—	—	—	
8	0.05	−0.01	0.07	0.25**	−0.19**	−0.18**	−0.18**	—
9	0.27**	0.63***	0.25***	0.11	0.06	0.01	−0.04	−0.06
10	−0.03	0.12	0.03	−0.10	0.18	0.10	0.16	−0.03
11	−0.06*	−0.13*	−0.05*	−0.03	0.00	−0.15*	0.02	0.40***

Note. *** $p < .001$; ** $p < .01$; * $p < .05$; 1 = Presence; 2 = Embodiment; 3 = Motivation; 4 = Cognitive engagement; 5 = Extraneous cognitive processing; 6 = Essential cognitive processing; 7 = Cognitive load; 8 = Learning achievement; 9 = VR condition; 10 = Time; 11 = Prior knowledge.

0.069) but revealed 11 nonsignificant paths (see Table 6). These 11 paths were evaluated by steps and one was retained. Specifically, path H2 (motivation increases cognitive engagement) was retained in the final model, because after the path H15 (presence increases cognitive engagement) was removed, the path H2 exhibited statistical significance ($\beta = 0.26, p = 0.030$). The reverse was also true—when path H2 was removed, the path H15 also showed a statistical significance ($\beta = 0.22$, $p = 0.015$). However, retaining path H2 instead of H15 in the final model was more theoretically interpretable. Motivation is regarded as a requirement of cognitive engagement in CATLM; motivation is a mediator but not a moderator when we assess the effect of presence on cognitive engagement. If learners held no interest in the topic, they would likely not engage despite experiencing the sense of being in a virtual world during the intervention. Figure 2 shows the simplified structural model. Due to the differences in

Table 6. Parameter Estimates for the Hypothesized Structural Model.

	From	To	Std. Coefficient	Support Hypothesis
H1	Cognitive load	Cognitive engagement	−0.64***	Yes
H2	Motivation	Cognitive engagement	0.18	No
H3	Motivation	Cognitive load	−0.37**	Yes
H4	Motivation	Learning achievement	−0.04	No
H5	Prior knowledge	Essential cognitive processing	−0.16*	Yes
H6	Prior knowledge	Learning achievement	0.41***	Yes
H7	Cognitive engagement	Learning achievement	0.28**	Yes
H8	VR format	Cognitive load	0.01	No
H9	VR format	Presence	<0.01	No
H10	VR format	Embodiment	0.65***	Yes
H11	Embodiment	Presence	0.46**	Yes
H12	Embodiment	Extraneous cognitive processing	0.16	No
H13	Presence	Motivation	0.30*	Yes
H14	Embodiment	Motivation	0.26*	Yes
H15	Presence	Cognitive engagement	0.14	No
H16	Embodiment	Cognitive engagement	<0.01	No
H17	Time	Motivation	0.05	No
H18	Time	Presence	−0.08	No
H19	Time	Embodiment	−0.06	No
H20	Time	Cognitive load	0.17	No

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

range among scales, these path coefficients were standardized. The model showed a good fit (χ^2/df = 1.347, RMSEA = 0.049, CFI = 0.924, TLI = 0.914, SRMR = 0.072).

In addition to the ten significant direct paths shown in Figure 2, VR format had a significant indirect effect on presence through embodiment ($\beta = 0.28$, $p < 0.001$). The test also provided the R^2 to indicate the amount of variance in the dependent factors that could be explained by other factors and external variables. Overall, the model explained 19.6% of the variance in presence, 39.2% in embodiment, 23.9% in motivation, 57.8% in cognitive engagement, 46.3% in essential processing, 100.0% in extraneous processing, 10.5% in cognitive load, and 22.6% in learning achievement. All of these values satisfied the cut-off value of R^2 ($R^2 \geq 10.0\%$) in SEM (Falk & Miller, 1992).

Model Modification

To further explore the simplified structural model, correlations between factors were examined based on the data in Table 5. Specifically, there were 11 statistically significant correlations that were not included in the initially hypothesized model. Thus, 11 potential new relations representing these correlations were added to the simplified structural model by steps. However, no significant relationships were observed after these elements were added to the model. Furthermore, we also explored the possible interaction between VR format and time. Results showed that this interaction had no significant effect on embodiment, presence, or cognitive load. Thus, the simplified structural model shown in Figure 2 was the final model in this study.

Discussion

Researchers have criticized that many VR learning studies have not been sufficiently guided by explicit learning theories, and there are few learning theories that incorporate distinctive elements associated with VR environments (Radianti et al., 2020). To advance the use of VR in effective instruction, this study proposed a model that

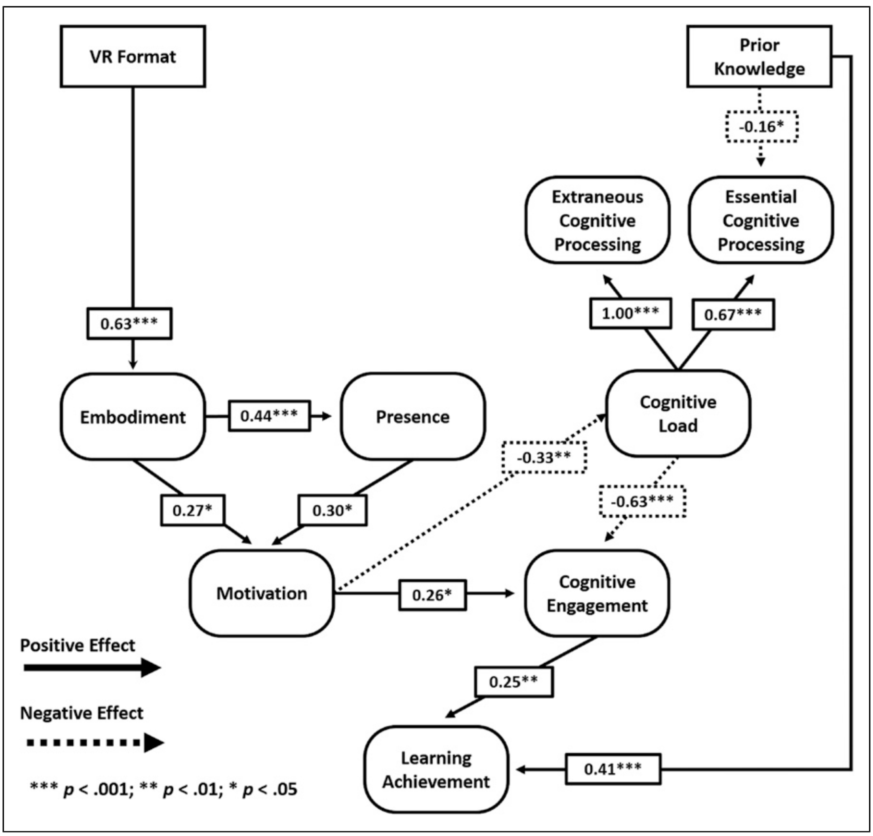

Figure 2. Simplified structural model.

extended CATLM into a VR learning context and evaluated this model using an SEM approach. Overall, the final SEM model (i.e., CATLM for Virtual Reality, or CATLM-VR) supported the core propositions of CATLM. This support demonstrated that the original CATLM was a relevant initial framework for VR studies and practice. In addition, the new model also illustrated how VR format impacts VR experiences and learning. This illustration addressed key gaps by incorporating specific VR factors in CATLM. To the best of our knowledge, this is the first study to theoretically extend CATLM with supporting empirical data. Using SEM, testing provided parameter estimates for all identified relationships among factors simultaneously with a consideration of measurement errors. Thus, the result appears reliable. Furthermore, the established model describes what future modules should consider including to become effective VR learning experiences, and thus has the potential to direct educators in the design VR learning applications and activities.

Are the principles of Cognitive-Affective Theory of Learning with Media applicable to Virtual Reality learning?

The first seven hypotheses included in the research model were directly derived from CATLM hypotheses. Among these seven hypotheses, following six were supported by the final SEM model:

H1: Cognitive load has a direct and negative effect on cognitive engagement.
H2: Motivation has a direct and positive effect on cognitive engagement.
H3: Motivation has a direct and negative effect on cognitive load.
H5: Prior knowledge has a direct and negative effect on essential cognitive processing.
H6: Prior knowledge has a direct and positive effect on learning achievement.
H7: Cognitive engagement has a direct and positive effect on learning achievement.

These six validated relationships encompass fundamental cognitive processes described in CATLM. First, the two-factor structure of cognitive load and the negative effect of cognitive load on cognitive engagement match the claim in CATLM that there are three types of cognitive processing, and cognitive processing is restricted by the learner's cognitive capacity. Second, the direct effect of prior knowledge on essential cognitive processing is aligned with the claim that prior knowledge can help the learner select meaningful new information in working memory. Third, the direct effects of motivation on cognitive load and cognitive engagement are consistent with the statement that motivation affects cognitive processing in learning. Fourth, the direct effect of cognitive load on cognitive engagement, and the direct effects of cognitive engagement and prior knowledge on learning achievement, support the described process that new information is selected, organized, and then integrated into new knowledge with existing knowledge.

Only H4 ("Motivation has a direct and positive effect on learning achievement") was not supported in the final model and therefore removed. This removal might be explainable with respect to the distracting effects of hedonic motivation during knowledge construction (Makransky et al., 2019)—if learners focus too much on enjoyment or "fun," they invest less effort in learning tasks. Therefore, there is not a straightforwardly positive effect of motivation on learning achievement. This removal is also aligned with the assumption in CATLM that cognitive engagement mediates the effect of motivation on learning achievement (i.e., learning is the result of cognitive processing of the information rather than merely desire to learn). Thus, it is rational to conclude that principles of CATLM are still applicable to VR learning based on the results of the first seven hypotheses.

How Does the Virtual Reality Format Impact Learning?

In contrast to initial expectations, VR format only showed a significant direct effect on embodiment, but not on presence or cognitive load. The weak relationship between VR format and cognitive load might be attributed to the relative ease of understanding the learning content in the VR application. Table 3 indicates that all items measuring perceived cognitive processing had a mean of less than 40 on the 0–100 scale. Thus, participants did not seem to perceive the learning process as highly demanding. This explanation is aligned with the claim that modality effects on learning are less applicable in low cognitively demanding conditions (Moreno, 2006; Sweller et al., 1998; Tindall-ford et al., 1997). Another possible explanation is that there was a mismatch between the advantage of gestures in learning and the questions in the knowledge quizzes. According to the cube of educational embodiment in VR (Johnson-Glenberg, 2018), manipulation helps individuals learn the content congruent with their gestures because of reduced cognitive demands. Particularly, in Johnson-Glenberg et al. (2021) recent path analysis on low and high embodied groups in either VR or on a desktop, the high embodied groups (with more congruent manipulation) always learned more content. In this study, grabbing and rotating virtual planets might help these participants remember the location of a site on a planet (e.g., a large canyon is on the north of a high mountain on Mars). However, the knowledge quizzes did not explicitly include questions that needed participants to indicate the position of a site. Thus, participants might not pay enough attention to this type of information in their VR learning experience, and the impact of VR format on cognitive load was small. The finding that VR format had a significant direct effect on embodiment, and a significant indirect effect on presence through embodiment, may be explainable if sense of presence is determined by visual display and interaction, simultaneously (Makransky & Petersen, 2021). The higher-immersion VR format did not possess a much more advanced visual display compared to the moderate immersion VR format, but did include substantive interaction affordances (e.g., bimanual control and body tracking). Thus, VR format had a significant effect on embodiment, and the effect of VR format on presence was mediated by embodiment in this model.

Furthermore, both presence and embodiment had a significant effect on motivation but not on cognitive engagement. These findings may be explained in terms of deeper perception of the surrounding environment and virtual body as antecedents—but not requirements—of cognitive engagement in multimodal learning (Hoffman & Nadelson, 2010; Hoffman & Novak, 2009). Prior studies that found the influence of presence and embodiment on cognitive engagement did not necessarily involve a consideration of motivation as the mediator (e.g., Animesh et al., 2011; Nijs et al., 2012; Zaman et al., 2010).

It is surprising that no significant relationship was found between embodiment and extraneous cognitive processing. This result may be explained from embodiment itself. A high level of embodiment implies a more flexible control of the virtual body and thus, more autonomy in a VR environment (Kilteni et al., 2012). Although cognitive processing can benefit from the ability of flexible controlling the virtual body, autonomy may lead to more unnecessary actions/exploration. These unnecessary actions can cause additional extraneous cognitive processing. Thus, the relationship between embodiment and extraneous cognitive processing is uncertain. This finding and explanation are aligned with Makransky et al. (2020) study in which students in the VR group performed not better than the video group until an optimal learning strategy was introduced in the VR group.

Finally, the factor of *time* (i.e., the number of sessions that a participant experienced in the VR environment) was not retained in the final model. Neither time nor its interaction with VR format exhibited any significant effect on other factors (i.e., motivation, embodiment, presence, and cognitive load). However, previous findings that students' motivation, and learning decreased over time in digital learning environments (Li & Ma, 2010; Tsay & Kofinas, 2018). This difference may be explained by the duration of the assessed learning activities. Both Li and Ma (2010) and Tsay and Kofinas' (2018) studies took one semester as the unit to measure students' changes. Our study only included three sessions and lasted approximately half a month for each participant. In other words, longer duration may be an important factor when we assess the impact of time (i.e., novelty) on motivation and learning.

In conclusion, evaluations of the extended model indicated that the impact of VR format on learning derived primarily from its influence on learners' motivation in this study. Due to the relatively large difference in behavior control affordances, the direct effect of VR format on learners' senses stemmed more from sense of embodiment than presence. Nevertheless, both embodiment and presence increased learners' motivation. A high level of motivation fostered cognitive engagement while a high level of cognitive engagement seemed to support more effective learning in VR.

Limitations

It is worth mentioning that there are potential limitations to this study. Data used for the model evaluation were collected from participants three times. A correlation likely exists between data points from the same participant in different sessions. Thus, we conducted exploratory factor analyses for data in each session. Overall, the factor structure in our SEM was stable in each session (see Supplemental Material). In other words, the possible bias caused by the repeated measure is small.

Second, the factors in CATLM-VR are still limited. Beyond the constructs that our study included, the original CATLM also mentions other factors such as metacognitive skills, emotion, and feedback. Additionally, VR learning may be affected by some factors not included in both CATLM and CATLM-VR, such as learners' previous VR experience, interest, and the congruence of gestures to the learning content. A future consideration is how to best integrate these factors and develop a more comprehensive model that enriches our understanding. Third, this study only used one VR application (with two versions) to collect data from adult undergraduate students at one university in the United States. Participants felt that the learning content in this VR application was easily understood, but manipulation was not quite useful because they could get most of the knowledge from the textual reading. However, learners who have less prior knowledge (e.g., middle school students) may not feel the VR learning content is simple and easily understood. Manipulation may be more effective to help these novices than adult undergraduates (Kalyuga, 2007). In this situation, VR format may have a significant effect on cognitive load. Thus, a more generalized model is advocated to be built in the future based on more factors, different types of VR applications, and a broader population.

Implications

This study has theoretical and practical implications for VR learning in motivational and cognitive aspects, respectively. VR can afford learners a sense of embodiment and presence that contribute to higher motivation. The motivational path, which proceeds from VR format to motivation, confirms the value of using multi-processing channels on enhancing learners' user experience in a VR environment (Anazco, 2020; Birt et al., 2015; Deng, 2017; Fernandes et al., 2016; Shu et al., 2019; Ta, 2018). Some researchers claimed that users' increased motivation in a new media environment was due to the novelty and this increasing was unsustainable (Clark, 1983; Koch et al., 2018). CATLM-VR reveals that the impact of time (i.e., novelty) on user experience, motivation, and cognitive load were small and ignorable in this study. In other words, the benefits of high immersion on user experience and motivation are inherent. Increasing the level of immersion is an effective method to provide better user experience to learners and enhance their motivation in VR instruction.

Although a higher level of immersion can enhance motivation, the effect of immersion—or more broadly, multimodal interaction—on learning is uncertain. High immersion has the potential to facilitate learning (Makransky & Petersen, 2021). However, CATLM-VR reveals that VR format did not have a significant effect on cognitive load for the learning content was simple in our study. This finding can explain why additional multi-sensory stimuli in a VR environment can neither decrease cognitive load nor benefit learning achievement in some situations (i.e., Anazco, 2020;

Deng, 2017). Therefore, if the learning content is easily understood, educators may not need to select higher-immersion (and more expensive) formats for their VR instruction. In addition, motivation and cognitive engagement play mediation roles in a VR learning process. Enhancing students' motivation and cognitive engagement may more directly increase learning achievement than increasing the level of immersion and may be more universally applicable in VR instruction. The specific methods can be encouragement, positive feedback (Deci & Ryan, 2010), and designing activities that help students construct and elevate knowledge by themselves (e.g., reflection, self-explain, and debate) (Chi & Wylie, 2014).

Conclusion

This study proposed a model to extend CATLM into VR learning contexts, which was evaluated via SEM methods. The final, simplified CATLM-VR model supported the core principles and assumptions of CATLM in a VR context and addressed key gaps by incorporating specific VR-related factors in CATLM. To the best of our knowledge, this is the first study to theoretically extend CATLM based on supporting empirical data. Findings in this study have the potential to guide the design of VR learning applications and VR learning activities.

Acknowledgments

The authors would like to thank Dr Robert Gray, Zhen Zhao, Elena Kalina, Vanessa Ly, Ricardo Nieland Zavela, Samuel Arnold, and Chelsea Johnson for their assistance.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially funded by the Graduate Research Support Program (GRSP) at Arizona State University.

ORCID iDs

Wen Huang https://orcid.org/0000-0002-9155-0547
Scotty D. Craig https://orcid.org/0000-0002-4103-5547
Mina C. Johnson-Glenberg https://orcid.org/0000-0003-0941-3744

Supplemental Material

Supplemental material for this article is available online.

References

Adadan, E. (2020). Analyzing the role of metacognitive awareness in preservice chemistry teachers' understanding of gas behavior in a multirepresentational instruction setting. *Journal of Research in Science Teaching*, *57*(2), 253–278. https://doi.org/10.1002/tea.21589

Altas, B. (2015). Knowledge construction in multimedia learning at both cognitive and affective level. *Procedia—Social and Behavioral Sciences*, *191*, 1448–1454. https://doi.org/10.1016/j.sbspro.2015.04.709

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review*, *40*(1), 37–47. https://doi.org/10.2307/2094445

Anazco, M. S. (2020). *Multimodal virtual learning environments: The effects of visuo-haptic simulations on conceptual learning* [Doctoral dissertation, Purdue University].

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423.

Angulo, A., & Velasco, G. V. D. (2014). Immersive simulation of architectural spatial experiences (Eds.), Proceedings of the XVII Conference of the Iberoamerican Society of Digital Graphics (SIGraDi) (pp. 495-499). SIGraDi.

Animesh, A., Pinsonneault, A., Yang, S., Oh, W., Animesh, A., & Pinsonneauit, A. (2011). An odyssey into virtual worlds: Exploring the impacts of technological and spatial environments on intention to purchase virtual products. *Mis Quarterly*, *35*(3), 789–810.

Aukstakalnis, S., & Blatner, D. (1992). *Silicon mirage: The art and science of virtual reality*. Peach Pit Press.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*(5), 389–400. https://doi.org/10.1016/j.learninstruc.2006.09.001

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74-94. https://doi.org/10.1007/BF02723327

Barab, S., Zuiker, S., Warren, S., Hickey, D. A. N., Ingram-goble, A., Kwon, E., Kouper, I., & Herring, S. C. (2007). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education*, *90*(6), 1073–1091. https://doi.org/10.1002/sce

Bennie, S. J., Ranaghan, K. E., Deeks, H., Goldsmith, H. E., O'Connor, M. B., Mulholland, A. J., & Glowacki, D. R. (2019). Teaching enzyme catalysis using interactive molecular dynamics in virtual reality. *Journal of Chemical Education*, *96*(11), 2488–2496. https://doi.org/10.1021/acs.jchemed.9b00181

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Multivariate Software, Inc.

Bianchi-Berthouze, N. (2013). Understanding the role of body movement in player engagement. *Human–Computer Interaction*, *28*(1), 40–75. https://doi.org/10.1080/07370024.2012.688468

Biard, N., Cojean, S., & Jamet, E. (2018). Effects of segmentation and pacing on procedural learning by video. *Computers in Human Behavior*, *89*, 411-417. https://doi.org/10.1016/j.chb.2017.12.002.

Birt, J., Hovorka, D. S., & Nelson, J. (2015). Interdisciplinary translation of comparative visualization. ACIS 2015 Proceedings—26th Australasian Conference on Information Systems, November 30–December 4, 2015, Adelaide, Australia.

Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior*, *28*(3), 771–780. https://doi.org/10.1016/j.chb.2011.11.020

Brancheau, J. C. (1989). *The diffusion of information technology: Testing and extending innovation diffusion theory in the context of end-user computing* [Doctoral dissertation, University of Minnesota].

Brondi, R., Avveduto, G., & Carrozzino, M. (2015). Immersive technologies and natural interaction to improve serious games engagement (Eds.), International Conference on Games and Learning Alliance.https://doi.org/10.1007/978-3-319-40216-1.

Brown, I. T. J. (2004). Testing and extending theory in strategic information systems planning through literature analysis. *Information Resources Management Journal*, *17*(4), 20–48. https://doi.org/10.4018/irmj.2004100102

Bystrom, K.-E., Barfield, W., & Hendrix, C. (1999). A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *8*(2), 241–244

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., & Lancaster, M. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, *42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chukwudi, I., Zhang, M., & Gable, G. (2020). Extensive theory testing using case study. 40th International Conference on Information Systems, ICIS 2019 (pp. 1−17. Bitektine.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, *25*(2), 315–324.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, *53*(4), 445–459.

Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the academy of management journal. *Academy of Management Journal*, *50*(6), 1281–1303. https://doi.org/10.5465/AMJ.2007.28165855.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*(4), 715–730.

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement in everyday life*. Basic Books.

Deci, E. L., & Ryan, R. M. (1994). *Intrinsic motivation inventory*. Center for Self-Determination Theory.

Deci, E. L., & Ryan, R. M. (2010). Intrinsic motivation. In I. B. Weiner, & W. E. Craighead (Eds.), *The corsini encyclopedia of psychology* (pp. 1-2). John Wiley & Sons, Inc.

Deng, S. (2017). *Multimodal interactions in virtual environments using eye tracking and gesture control*. British Library EThOS service. http://eprints.bournemouth.ac.uk/30470/.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, *2*(2), 119–143.

Domagk, S., Schwartz, R. N., & Plass, J. L. (2010). Interactivity in multimedia learning: An integrated model. *Computers in Human Behavior*, *26*(5), 1024–1033. https://doi.org/10.1016/j.chb.2010.03.003

DrashVR (2019). *Titans of Space*. http://www.drashvr.com/.

Engelmann, J. B., & Pessoa, L. (2007). Motivation sharpens exogenous spatial attention. *Emotion*, *7*(3), 668–674. https://doi.org/10.1037/1528-3542.7.3.668

Falk, R., & Miller, N. B. (1992). *A primer for soft modeling*. University of Akron Press.

Fernandes, L. M. A., Cruz Matos, G., Azevedo, D., Rodrigues Nunes, R., Paredes, H., & Morgado, L. (2016). Exploring educational immersive videogames: An empirical study with a 3D multimodal interaction prototype. *Behaviour & Information Technology*, *35*(11), 907–918. https://doi.org/10.1080/0144929X.2016.1232754

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, *33*(4), 441–467. https://doi.org/10.1177/1046878102238607

Goldin-Meadow, S. (2011). Learning through gesture. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 595–607. https://doi.org/10.1002/wcs.132

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516–522

Gonzalez-Franco, M., & Peck, T. C. (2018). Avatar embodiment: Towards a standardized questionnaire. *Frontiers in Robotics and AI*, *5*, 1-9. https://doi.org/10.3389/frobt.2018.00074

Grabowski, A. (2019). Innovative and comprehensive support system for training people working in dangerous conditions. International Conference on Human-Computer Interaction, Orlando, FL, July 26–31, 2019 (pp. 394-405). https://doi.org/10.1007/978-3-030-22216-1

Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology*, *50*(5), 2572–2593. https://doi.org/10.1111/bjet.12864

Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Publications.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Prentice Hall, Inc.

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, *12*(1), 159.

Hoffman, B., & Nadelson, L. (2010). Motivational engagement and video gaming: A mixed methods study. *Educational Technology Research and Development*, *58*(3), 245–270. https://doi.org/10.1007/s11423-009-9134-9

Hoffman, D. L., & Novak, T. P. (2009). Flow online: Lessons learned and future prospects. *Journal of Interactive Marketing*, *23*(1), 23–34. https://doi.org/10.1016/j.intmar.2008.10.003

Hogan, N. S., & Schmidt, L. A. (2002). Testing the grief to personal growth model using structural equation modeling. *Death Studies*, *26*(8), 615–634. https://doi.org/10.1080/07481180290088338

Holmes, A. G. (2018). The role of interest and enjoyment in determining students' approach to learning. *Educational Process: International Journal*, *7*(2), 140–150. https://doi.org/10.22521/edupij.2018.72.4

Hou, H.-T., & Li, M.-C. (2014). Evaluating multiple aspects of a digital educational problem-solving-based adventure game. *Computers in Human Behavior*, *30*, 29–38. https://doi.org/10.1016/j.chb.2013.07.052

Hsu, C. Y., Tsai, C. C., & Wang, H. Y. (2012). Facilitating third graders' acquisition of scientific concepts through digital game-based learning: The effects of self-explanation principles. *The Asia-Pacific Education Researcher*, *21*(1), 71–82.

Huang, W. (2019). Examining the impact of head-mounted display virtual reality on the science self-efficacy of high schoolers. *Interactive Learning Environments*, *4*, 1-13. https://doi.org/10.1080/10494820.2019.1641525

Huang, W. (2020). *Investigating the novelty effect in virtual reality on STEM learning* [Doctoral dissertation, Arizona State University].

Huang, W., & Roscoe, R. D. (2021). Head-mounted display-based virtual reality systems in engineering education: A review of recent research. *Computer Applications in Engineering Education*, *29*(5), 1420–1435. https://doi.org/10.1002/cae.22393

Huang, W., Roscoe, R. D., Johnson-Glenberg, M. C., & Craig, S. D. (2021). Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *Journal of Computer Assisted Learning*, *37*(3), 745–758. https://doi.org/10.1111/jcal.12520

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 37–41. https://doi.org/10.1080/10705519909540118

Hu, F.-T., Ginns, P., & Bobis, J. (2015). Getting the point: Tracing worked examples enhances learning. *Learning and Instruction*, *35*, 85-93. https://doi.org/10.1016/j.learninstruc.2014.10.002

Jaccard, J., & Jacoby, J. (2010). *Theory construction and model-building skills: A practical guide for social scientists* (1st ed.). Guilford Publications.

Jackson, S., & Eklund, B. (2004). *The flow scales manual*. Publishers Graphics.

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, *5*(81), 81–19. https://doi.org/10.3389/frobt.2018.00081.

Johnson, C. I., Bailey, S. K. T., & Buskirk, W. L. V. (2017). Designing effective feedback messages in serious games and simulations: A research review. In H. Oostendorp & P. Wouters. *Instructional techniques to facilitate Learning and motivation of serious Games* (pp. 119-140). Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1

Johnson-Glenberg, M. C., Bartolomea, H., & Kalina, E. (2021). Platform is not destiny: Embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. *Journal of Computer Assisted Learning*, *37*(5), 1263–1284. https://doi.org/10.1111/jcal.12567

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(3), 23-31.

Keller, J. M. (Eds.). (2010). Generating and sustaining attention forethought. *Motivational design for learning and performance: The ARCS model approach* (pp. 75-96). Springer US. https://doi.org/10.1007/978-1-4419-1250-3

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, *21*(4), 103–105. https://doi.org/10.1162/PRES_a_00124

Koch, M., Luck, K. V., Schwarzer, J., & Draheim, S. (2018). The novelty effect in large display deployments—Experiences and lessons-learned for evaluating prototypes (Eds.), IProceedings of 16th European Conference on computer-supported Cooperative work-exploratory Papers, Salzburg, Austria, June 8-12, 2019. https://doi.org/10.18420/ecscw2018.

Kokkinara, E., Kilteni, K., Blom, K. J., & Slater, M. (2016). First person perspective of seated participants over a walking virtual body leads to illusory agency over the walking. *Scientific Reports*, *6*(1), 28879–28911. https://doi.org/10.1038/srep28879

Krämer, N. C., & Bente, G. (2010). Personalizing e-learning. The social effects of pedagogical agents. *Educational Psychology Review*, *22*(1), 71–87. https://doi.org/10.1007/s10648-010-9123-x

Kuldas, S., Satyen, L., Ismail, H. N., & Hashim, S. (2014). Greater cognitive effort for better learning: Tailoring an instructional design for learners with different levels of knowledge and motivation. *Psychologica Belgica*, *54*(4), 350–373. https://doi.org/10.5334/pb.aw

Lee, Y.-H., & Heeter, C. (2017). The effects of cognitive capacity and gaming expertise on attention and comprehension. *Journal of Computer Assisted Learning*, *33*(5), 473–485. https://doi.org/10.1111/jcal.12193

Lee, E. A.-L., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education*, *55*(4), 1424–1442. https://doi.org/10.1016/j.compedu.2010.06.006

Lepper, M. R., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology*, *97*(2), 184–196. https://doi.org/10.1037/0022-0663.97.2.184

Leppink, J., Paas, F., & Van Der Vleuten, C. P. M. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Leutner, D. (2014). Motivation and emotion as mediators in multimedia learning. *Learning and Instruction*, *29*, 174-175. https://doi.org/10.1016/j.learninstruc.2013.05.004.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, *22*(3), 215–243. https://doi.org/10.1007/s10648-010-9125-8

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education*, *95*(1), 174-187. https://doi.org/10.1016/j.compedu.2016.01.001

Littlejohn, S. W., Foss, K. A., & Oetzel, J. G. (2017). *Theories of human communication* (11th ed.). Waveland Press, Inc.

Løkke, A., & Sørensen, P. D. (2014). Theory testing using case studies. *The Electronic Journal of Business Research Methods*, *12*(1), 66-74.

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, *113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education (pp. 1-24). Educational Technology EResearch and Development.

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, *33*(1), 937–958. https://doi.org/10.1007/s10648-020-09586-2

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*(1), 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Malone, T., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction: Vol. III. Cognitive and affective process analysis* (pp. 223–253). Lawrence Erlbaum.

March, R., & Woodside, A. G. (2005). Testing theory of planned versus realized tourism behavior. *Annals of Tourism Research*, *32*(4), 905–924. https://doi.org/10.1016/j.annals.2004.07.012

Mashfufah, A., Nurkamto, J., Sajidan, W., & Novenda, I. L. (2019). Conceptual: Digital book in the era of digital learning approaches (DLA). *IOP Conference Series: Earth and Environmental Science*, *243*(1), 12107. https://doi.org/10.1088/1755-1315/243/1/012107

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction*, *29*, 171–173. https://doi.org/10.1016/j.learninstruc.2013.04.003

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, *33*(5), 403–423. https://doi.org/10.1111/jcal.12197

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52. https://doi.org/10.1207/S15326985EP3801

McConnel, J. (2018). "Why are we doing this?": Using digital reflection to increase student engagement. *Ubiquitous Learning*, *11*(2), 13–22. https://doi.org/10.18848/1835-9795/CGP/v11i02/13-22

McGuinness, C. (1990). Talking about thinking: The role of metacognition in teaching thinking. In M. D. K. Gilhooly & G. Erdos (Eds.), *Lines of thinking* (pp. 310-312). Academic.

McGuire, S. Y. (2015). *Teach students how to learn: Strategies you can incorporate into any course to improve student metacognition, study skills, and motivation*. Stylus Publishing, LLC.

Mcmillan, K., Flood, K., & Glaeser, R. (2017). Virtual reality, augmented reality, mixed reality, and the marine conservation movement. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *27*(S1), 162–168. https://doi.org/10.1002/aqc.2820

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems*, *E77*(12), 1–15.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, *22*(3), 149–158. https://doi.org/10.1111/j.1365-2729.2006.00170.x

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

Nijs, L., Coussement, P., Moens, B., Amelinck, D., Lesaffre, M., & Leman, M. (2012). Interacting with the music paint machine: Relating the constructs of flow experience and presence. *Interacting With Computers*, *24*(4), 237–250. https://doi.org/10.1016/j.intcom.2012.05.002

Onyesolu, M. O., & Eze, F. U. (2011). Understanding virtual reality technology: Advances and applications. In D. Zeng (Eds.), *Advances in computer science and Engineering*. InTech.

Oxford (2019). *Oxford Dictionaries*. Oxford University Press. https://www.oxforddictionaries.com/.

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

Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy context on children's learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology*, *62*(4), 625–633.

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

Petersen, G. B., Klingenberg, S., Mayer, R. E., & Makransky, G. (2020). The virtual field trip: Investigating how to optimize immersive virtual learning in climate change education. *British Journal of Educational Technology*, *51*(6), 2099–2115. https://doi.org/https://doi.org/10.1111/bjet.12991

Pineda, B. (2015). *Using computer simulations as a pre-training activity in a hands-on lab to help community college students improve their understanding of physics*. [Doctoral dissertation, The University of San Francisco].

Pintrich, P. R. (2003). Motivation and classroom learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (pp. 103–122). Wiley.

Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In M. Gartmeier & S. Y. Tettegah (Eds.), *Emotions, technology, design, and Learning* (pp. 131-161). Elsevier Inc. https://doi.org/10.1016/b978-0-12-801856-9.00007-4.

Plass, J. L., & Schwartz, R. N. (2014). Multimedia learning with simulations and microworlds. In R. E. Mayer (Ed.), *The cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.036.

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

Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & Mcmahan, R. P. (2015). Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Transactions on Visualization and Computer Graphics*, *21*(7), 794–807.

Ritterfeld, U., Shen, C., Wang, H., Nocera, L., & Wong, W. L. (2009). Multimodality and interactivity: Connecting properties of serious games with educational outcomes. *Cyberpsychology and Behavior*, *12*(6), 691–697. https://doi.org/10.1089/cpb.2009.0099

Robinson, L. J., Stevens, L. H., Threapleton, C. J., Vainiute, J., McAllister-Williams, R. H., & Gallagher, P. (2012). Effects of intrinsic and extrinsic motivation on attention and memory. *Acta Psychologica*, *141*(2), 243–249. https://doi.org/10.1016/j.actpsy.2012.05.012

Roets, A., Van Hiel, A., Cornelis, I., & Soetens, B. (2008). Determinants of task performance and invested effort: A need for closure by relative cognitive capacity interaction analysis. *Personality & Social Psychology Bulletin*, *34*(6), 779–792. https://doi.org/10.1177/0146167208315554

Sajjadi, P., Hoffmann, L., Cimiano, P., & Kopp, S. (2018). On the effect of a personality-driven ECA on perceived social presence and game experience in VR. 10th International Conference on virtual worlds and Games for serious applications, VS-Games 2018 - Proceedings, Würzburg, Germany, September 5–7 2018. https://doi.org/10.1109/VS-Games.2018.8493436.

Salzman, M. C., Dede, C., Loftin, R. B., & Chen, J. (1999). A model for understanding how virtual reality aids complex conceptual learning. *Presence: Teleoperators and Virtual Environments*, *8*(3), 293–316. https://doi.org/10.1162/105474699566242

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, *94*(3), 241–255. https://doi.org/10.1016/j.cognition.2004.04.003

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, *26*(3–4), 299–323. https://doi.org/10.1207/s15326985ep2603&4_5

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469–508. https://doi.org/10.1007/s10648-007-9053-4

Schrader, C., & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior*, *28*(2), 648–658. https://doi.org/10.1016/j.chb.2011.11.011

Shernoff, D. J., & Strati, A. D. (2011). Advances in engineering education measuring engagement as students learn dynamic systems and control with a video game department of mechanical engineering. *Advances in Engineering Education*, *2*(3), 1–32.

Shin, D. (2017). The role of affordance in the experience of virtual reality learning: Technological and affective affordances in virtual reality. *Telematics and Informatics*, *34*(8), 1826–1836. https://doi.org/10.1016/j.tele.2017.05.013

Shin, D. H., Biocca, F., & Choo, H. (2013). Exploring the user experience of three-dimensional virtual learning environments. *Behaviour and Information Technology*, *32*(2), 203–214. https://doi.org/10.1080/0144929X.2011.606334

Shu, B., Sziebig, G., & Pieters, R. (2019). Architecture for safe human-robot collaboration: Multi-modal communication in virtual reality for efficient task execution. *IEEE International Symposium on Industrial Electronics*, (pp. 2297–2302). https://doi.org/10.1109/ISIE.2019.8781372.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., McCarthy, J., & Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(3), 469–477.

Slater, M., Usoh, M., & Chrysanthou, Y. (1995). The influence of dynamic shadows on presence in immersive virtual environments. *Virtual Environments'*, *95*, 8–21. https://doi.org/10.1007/978-3-7091-9433-1.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in immersive virtual environments. *Presence: Teleoperators & Virtual Environments*, *3*(2), 130–144.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1007/s10750-008-9541-7

Song, J., Kim, S. I., & Bong, M. (2019). The more interest, the less effort cost perception and effort avoidance. *Frontiers in Psychology*, *10*, 1–13. https://doi.org/10.3389/fpsyg.2019.02146.

Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research*, *81*(1), 97–127. https://doi.org/10.3102/0034654310378174.

Spek, E. D. V. D., Wouters, P., & Oostendorp, H. V. (2008). Efficient learning in serious games: A cognition-based design guidelines approach.In. Conference Proceedings for the Association for Educational Communications and Technology Conference, 18 December, 2021 (pp. 382–390).

Straub, D., Limayem, M., & Karahanna-Evaristo, E. (1995). Measuring system usage: Implications for IS theory testing. *Management Science*, *41*(8), 1328–1342. https://doi.org/10.1287/mnsc.41.8.1328

Svinicki, M. D. (2004). Motivating students to learn.In. M. D. Svinicki (Ed.), *Learning and motivation in the postsecondary classroom* (pp. 141-176). Jossey-Bass.

Sweller, J. (1999). *Instructional design in technical areas*. ACER Press.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., Merrienboer, J. J. G. V., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/a

Ta, Z. (2018). *Proposing an instructional design model with designing and developing sensory immersive VRLE to teach concepts and procedures*. Academic Studies in Educational Science.

Takatalo, J. (2002). *Presence and flow in virtual environments: An explorative study*. University of Helsinki.

Tcha-tokey, K., Christmann, O., Loup-escande, E., Loup, G., & Richir, S. (2018). Towards a model of user experience in immersive virtual environments. *Advances in Human-Computer Interaction*, *2018*, 1-10.

Timcenko, O., Kofoed, L. B., Schoenau-fog, H., & Reng, L. (2017). Purposive game production in educational setup investigating team collaboration in virtual reality. *International Conference on Human-Computer Interaction*, *714*, 184-191. https://doi.org/10.1007/978-3-319-58753-0.

Tindall-ford, S., Chandler, P., Sweller, J., & Bartlett, S. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology. Applied*, *3*(4), 257–287.

Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, *34*(4), 414–424. https://doi.org/10.1037/0022-0167.34.4.414

Tsay, C. H.-H., & Kofinas, A. (2018). Student engagement and the novelty effect in a technology-mediated gamified course. *Academy of Management Proceedings*, *2018*(1), 13030. https://doi.org/10.5465/AMBPP.2018.13030abstract

Tulving, E. (1977). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic.

Valverde-Berrocoso, J., del Carmen Garrido-Arroyo, M., Burgos-Videla, C., & Morales-Cevallos, M. B. (2020). Trends in educational research about e-learning: A systematic literature review (2009-2018). *Sustainability*, *12*(12), 5153. https://doi.org/10.3390/su12125153

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and strucutural equation modeling. *Personnel Psychology*, *48*(4), 865–885.

Wade, S., & Kidd, C. (2019). The role of prior knowledge and curiosity in learning. *Psychonomic Bulletin & Review*, *26*(4), 1-11.

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons Inc.

Ward, P. R., Mamerow, L., & Meyer, S. B. (2014). Interpersonal trust across six asia-Pacific countries: Testing and extending the "high trust society" and "low trust society" theory. *Plos One*, *9*(4), e95555. https://doi.org/10.1371/journal.pone.0095555

Whitton, N. (2011). Game engagement theory and adult learning. *Simulation & Gaming*, *42*(5), 596–609. https://doi.org/10.1177/1046878110378587

Wicken, C. D. (1992). Introduction to engineering psychology and human performance. In C. D. Wicken (Ed.), *Engineering psychology and human performance* (pp. 1–23). Ebook.

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., & Biocca, F. (2007). A process model of the formation of spatial presence experiences. *Media Psychology*, *9*(3), 493–525. https://doi.org/10.1080/15213260701283079

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, *7*(3), 225–241.

Zaman, M., Ananda, M., & Dai, Q. (2010). Experiencing flow with instant messaging and its facilitating role on creative behaviors. *Computers in Human Behavior*, *26*(5), 1009–1018. https://doi.org/10.1016/j.chb.2010.03.001

Zestcott, C. A. (2017). *The role of conscious attention in embodiment: Initial evidence of a dual process model of embodied cognition*. The university of arizona.

Author Biography

Dr. Wen Huang is a Postdoctoral Fellow in the Department of Teaching Learning within the Simmons School of Education and Human Development at Southern Methodist University. He completed his PhD at Arizona State University in Engineering Education Systems and Design Program in 2020. His research focuses on using immersive technologies in learning and education. He is particularly interested in modeling and longitudinal evaluations.

Dr. Rod D. Roscoe is an Associate Professor of Human Systems Engineering in the Polytechnic School of the Ira A. Fulton Schools of Engineering, and a Diane and Gary Tooker Professor of Effective Education in STEM at Arizona State University, USA. He is affiliate faculty of the Mary Lou Fulton Teachers College; the Center for Gender Equity in Science and Technology (CGEST), and the Center for Human, Artificial Intelligence, and Robot Teaming (CHART). His research investigates how the intersection of learning science, computer science, and user science can inform effective and innovative uses of educational technologies. He is also interested in how cognitive, metacognitive, and motivational learning processes can be scaffolded by educational technology.

Dr. Scotty D. Craig is an Associate Professor of Human Systems Engineering at Arizona State University and Director of the DoD Advanced Distributed Learning Initiative Partnership Laboratory at Arizona State University (adl.asu.edu). He has a dual affiliation with the Ira A. Fulton Schools of Engineering and the Mary Lou Fulton Teachers College. Dr. Craig has expertise at the intersection of cognitive psychology, usability, and the science of learning (specifically learning technology). He has served on the program committee for AIED Society, iFEST, and Cognitive Science Society and the editorial board for Journal of Educational Psychology, Computers & Education Open, International Journal of STEM Education, and the journal Technology, Mind, & Behavior.

Dr. Mina C. Johnson-Glenberg is a Research Professor at Arizona State University in the Psychology department. She is also an entrepreneur and President of the award-winning learning technology company called Embodied Game, LLC. At her university lab and in the spinout company, she leads a team that creates, researches, and disseminates innovative virtual and mixed reality (XR) educational content for 4th graders through life-long learners. Several of the NSF-funded STEM games can be downloaded for free at www.embodied-games.com. She has a varied background as a screenwriter, experimental psychologist, and game designer. Her lab has been at the forefront of creating "embodied" and kinesthetically-active games using sensors and motion capture technologies for over a decade and a half. Dr. Johnson-Glenberg presents nationally and internationally. Some recent publications include design guidelines for learning VR at Frontiers in Robotics and AI and, a rubric that teachers can use to score the quality of a VR lesson.

Appendix A

Table A1. Subjective Experiences Questionnaire Items.

Factor	Item Text	Source
Presence	Please rate your sense of being in the universe. I had a sense of being in the universe… [0 = "at no time" to 100 = "almost all the time"]	Slater et al. (1994, 1995)
	To what extent were there times during the experience when the universe was the "reality" for you? There were times during the experience when the universe was the reality for me… [0 = "at no time" to 100 = "almost all the time"]	
	Do you think of the universe more as images that you saw, or more as somewhere that you visited? The universe seemed to be more like… [0 = "images that I saw" to 100 = "somewhere I visited"]	
	During the experience, did you often think to yourself that you were just in a room, or did the universe overwhelm you? During the experience I was thinking that I was really in the room… [0 = "most of the time" to 100 = "rarely"]	
Embodiment	[Items were rated on a scale of 0 = "strongly disagree" to 100 = "strongly agree"]	Gonzalez-Franco and Peck (2018)
	It felt as if the virtual body or body part I saw was someone else's. (R)	
	It felt like I could control the virtual body or body part as if it was my own	
	I felt as if the movements of the virtual body were influencing my own movements	
Motivation	[Items were rated on a scale of 0 = "not true at all" to 100 = "very true"]	Deci and Ryan (1994)
	I enjoyed doing this activity very much	
	This activity is fun to do	
	I would describe this activity as very interesting	

(continued)

Table A1. (continued)

Factor	Item Text	Source
Engagement	[Items were rated on a scale of 0 = "never" to 100 = "always"]	Jackson and Eklund. (2004)
	I felt that I was competent enough to meet the demands of the learning task	
	I had a strong sense of what I wanted to do	
	I had a good idea about how well I was doing while I was involved in the activity	
	I was completely focused on the learning task at hand	
	I had a feeling of total control over what I was doing	
Essential cognitive processing	[Items were rated on a scale of 0 = "not at all the case" to 100 = "completely the case"]	Ayres (2006) and Leppink et al. (2013)
	The learning content was difficult	
	The learning content was complex	
Extraneous cognitive processing	[Items were rated on a scale of 0 = "not at all the case" to 100 = "completely the case"]	
	It was difficult to learn in this VR environment	Cierniak et al. (2009)