Title
The Impact of Individual Differences on Learning with Distractions

Author
Lawson, Alyssa

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

The Impact of Individual Differences on Learning with Distractions

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychological and Brain Sciences

by

Alyssa Pualani Lawson

Committee in charge:

Professor Richard E. Mayer, Chair

Professor Barry Giesbrecht

Professor Mary Hegarty

Professor Vanessa Woods, Lecturer

June 2023

The dissertation of Alyssa Pualani Lawson is approved.

Barry Giesbrecht

Mary Hegarty

Vanessa Woods

Richard E. Mayer, Committee Chair

May 2023

The Impact of Individual Differences on Learning with Distractions

ACKNOWLEDGEMENTS

To my family: Thank you for always believing in me and supporting my pursuit of a PhD. I know you did not always understand what I was doing, but your never-ending support in all areas of my life helped me get here.

To Zoom: Thank you for being with me on this journey every step of the way. And thank you for sleeping through all my Zoom classes during the pandemic.

To Will: Thank you for being the best cheerleader anyone could ask for along the way. Your love during the difficult times in this program is one of the reasons I am here today!

To Rich: Thank you for chatting with me at UCLA so long along and letting me nerd out about my research. Then still accepting me into your lab. Your mentorship has made me grow into a confident and knowledgeable researcher and person.

To Ashleigh: Thank you for being my lab big sister. I solved so many problems, research related and not, due to your advice and knowledge.

To my cohort: Thank you for being with me on this journey every step of the way. As Elle Woods would say, "WE DID IT!"

To my lab mates: Katie, Jocelyn, Fangzheng, Cynthia, Miriam, and Amedee: Thank you for all your support (and commiseration) at the different stages of grad school.

To my research assistants: I would not be graduating today without your dedication to data collection. Thank you so much for all your hard work.

To my committee: Thank you for your feedback, support, and wisdom regarding this dissertation and beyond. I cherish your support throughout this program.

VITA OF ALYSSA PUALANI LAWSON
April 2023

EDUCATION

Bachelor of Arts in Psychology, Chapman University, May 2016 (summa cum laude)
Master of Arts in Psychology, California State University, Los Angeles, May 2018
Doctor of Philosophy in Psychological and Brain Sciences, University of California, Santa Barbara, June 2023 (expected)

PROFESSIONAL EMPLOYMENT

2015-16: Undergraduate Research Assistant, Chapman University
2016-18: Student Research Assistant, California State University, Los Angeles
2018-20: Teaching Assistant, University of California, Santa Barbara
2021: Educational Testing Services (ETS) Summer Intern, University of California, Santa Barbara
2022: Summer Teaching Institute for Associates (STIA) Facilitator, University of California, Santa Barbara
2021-23: Teaching Assistant Pedagogical Advisor, University of California, Santa Barbara
2021-22: Teaching Associate, University of California, Santa Barbara
2022-23: Certificate in College and University Teaching Graduate Student Coordinator, University of California, Santa Barbara

PUBLICATIONS

Lawson, A. P., Mirinjian, A. & Son, J. Y. (2018). Preventing calculations helps students learn mathematics. *Journal of Cognitive Education and Psychology, 17*(2), 178-197.
Lawson, A. P., Davis, C., & Son, J. Y. (2019). Not all flipped classrooms are the same: Using learning science to design flipped classrooms. *Journal of the Scholarship of Teaching and Learning, 19*(5), 77-104. https://doi.org/10.14434/josotl.v19i5.25856
Lawson, A. P., & Mayer, R. E. (2021). The power of voice to convey emotion in multimedia instructional messages. *International Journal of Artificial Intelligence in Education*, *32*(4), 971-990. https://doi.org/10.1007/s40593-021-00282-y
Lawson, A. P., & Demeke, E. (2021). Developmental mathematics students' anxiety, mindset, and performance in algebra. *MathAMATYC, 13*(1), 46-53.
Lawson, A. P., Ayala, B., Son, J. Y. (2021). Priming students to calculate inhibits sense-making. *Journal of Cognitive Science, 22*(1), 41-70. DOI: 10.17791/jcs.2021/22/1/41
Lawson, A. P., & Mayer, R. E. (2021). Benefits of writing an explanation during pauses in multimedia lessons. *Educational Psychology Review, 33*, 1859-1885. https://doi.org/10.1007/s10648-021-09594-w
Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021). Do learners recognize and relate to the emotions displayed by virtual instructors? *International Journal of Artificial Intelligence in Education*, *31*, 134-153. https://doi.org/10/1007/s40593-021-00238-2

Adamo, N., Benes, B., Mayer, R. E., Lei, X., Wang, Z., Meyer, Z., & **Lawson, A.** (2021). Multimodal affective pedagogical agents for different types of learners. In D Russo et al. (Eds.), *Intelligent human systems integration*. (AISC 1322; pp. 218-224). Springer. https://doi.org/10.1007/978-3-030-68017-6_33

Lei, X., Adamo-Villani, N., Benes, B., Wang, Z., Meyer, Z., Mayer, R., & **Lawson, A**. (2021). Perceived naturalness of interpolation methods for character upper body animation. In G. Bebis et al. (Eds.), *Advances in visual computing* (LNCS 13017; pp. 103-115). Springer. https://doi.org/10.1007/978-3-030-90439-5_9

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021). Recognizing the emotional state of human and virtual instructors. *Computers in Human Behavior, 114*, 106554. https://doi.org/10.1016/j.chb.2020.106554

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Bedrich, B., Lei, X., & Cheng, J. (2021). The Positivity Principle: Do positive instructors improve learning from video lectures? *Educational Technology: Research and Development, 69*, 3101-3129. https://doi.org/10.1007/s11423-021-10057-w

Lawson, A. P., & Mayer, R. E. (2022). Does the emotional stance of human and virtual instructors in instructional videos affect learning processes and outcomes? *Contemporary Educational Psychology, 70*, 102080. https://doi.org/10.1016/j.cedpsych.2022.102080

Pataranutaporn, P., Leong, J., Danry, V., **Lawson, A. P**., Maes, P., & Sra, M. (2022). AI-generated virtual instructors based on liked or admired people can improve motivation and foster positive emotions for learning. *2022 IEEE Frontiers in Education Conference*, 1-9, doi: 10.1109/FIE56618.2022.9962478

AWARDS

2020-2023 National Defense Science and Engineering Graduate Fellowship, U. S. Department of Defense

2021 Summer Teaching Institute for Associates Certificate, University of California, Santa Barbara

2022 Grad Slam Runner Up, University of California, Santa Barbara

2022 Dixon-Levy GSA Service Award, Honorable Mention, University of California, Santa Barbara

2022 Goodchild Graduate Mentoring Award, University of California, Santa Barbara

2022 2nd Place in 2nd Annual Pitch Perfect Competition at Association for Psychological Science Annual Convention

2022 Graduate Conference Award, Psychonomic Society

FIELDS OF STUDY

Major Field: Cognitive and Educational Psychology

Studies in Human-Computer Interaction

Studies in Educational Technology

ABSTRACT

The Impact of Individual Differences on Learning with Distractions

by

Alyssa Pualani Lawson

Learning in a multimedia environment puts many demands on a learner's limited working memory, but this can become even more demanding as the level of distraction increases in a lesson. What has not been investigated much in previous literature is whether higher levels of distraction in lessons are more harmful to some learners than others. This series of studies investigates how individual differences in the ability to manage incoming information (i.e., executive function) and the ability to hold incoming information (i.e., working memory capacity) play a role in learning across lessons with various amounts and types of distractions. Experiment 1 investigated the role that individual differences in managing and holding incoming information play in learning from online multimedia lessons with various amounts of verbal and visual distracting elements. In Study 1a, learners watched a lesson with a high degree of distractions, in Study 1b learners watch a lesson with a moderate degree of distractions, and in Study 1c learners watched a lesson with a low degree of distractions. This set of studies found that as the degree of distraction increased, the correlation between executive function and posttest performance increased; there was no significant relationship at low levels of distraction but there was a significant relationship at moderate and high levels of distraction, suggesting that better executive function was related to better posttest performance only when a lesson was distracting. Working memory capacity was not significantly correlated with posttest performance in any version of the lesson. Experiment 2 further investigated this relationship to see if this finding would be extended into media with an increased possibility of distraction, specifically immersive virtual reality (IVR). In Experiment 2, learners watched a lesson presented via IVR or as a slideshow. This experiment found that, although there were no differences in learning between the two conditions, the relationship between executive function and posttest performance was only seen in the IVR lesson, not the slideshow lesson, such that learners with better executive function had better posttest scores when learning in the IVR lesson than learners with worse executive function. Once again, working memory capacity did not significant correlate with posttest performance. Lastly, Experiment 3 investigated whether the novelty of the IVR learning environment increased the distractibility of learning in IVR. Experiment 3 had learners either play a game in IVR prior to the lesson or had participants only watch the lesson. This experiment found that reducing the novelty of IVR for learners did not impact learning and so novelty may not be the reason for increased distractibility in IVR lessons. This series of experiments suggests that it is vital to think about the role that individual differences in learners' skill at managing incoming information plays in learning, especially when using lessons with distracting information or distracting technology.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter I: Introduction

Objectives and Rationale

Learning is a cognitively demanding task, as learners are asked to process a lot of information at one time. This activity can be made even more difficult by presenting instructional material containing distractions, defined as information that is presented in a lesson but is not relevant to the learning objectives of the lesson. For learning to occur, information must be processed through one's limited working memory before it can be moved into long-term memory and stored for future use (Atkinson & Shiffrin, 1968). However, if the learner is confronted with distractions in the lesson, they must either sort the relevant information from the irrelevant and process only what is necessary, or they must have the capacity to process all of the presented information, both relevant and irrelevant, at one time. This means that, in the presence of distractions, a learner's ability to manage incoming information and/or capacity to hold new information can play a role in their longer-term understanding of the lesson.

Furthermore, not everyone has the same cognitive tools for managing and holding information in working memory while learning new material. This may affect individuals' ability to learn from lessons, especially lessons that have an increased amount of distraction (e.g., Ackerman & Lohman, 2006; Carroll & Maxwell, 1979). For this reason, this series of studies focuses on how individual differences in these two areas of working memory–management and holding of new incoming information–can impact how well someone can learn from a distracting lesson or from using technology that may be inherently distracting, such as with immersive virtual reality.

The goal of this series of experiments is to understand how individual differences in managing and holding incoming information impact learning from distracting lessons and technologies. The main aim of this work is to answer the question of whether individual differences in learners' ability to manage information (i.e., executive function) and hold information (i.e., working memory capacity) are related to learning from distracting multimedia lessons and learning with immersive technology.

To investigate this question, the present studies operationalized the idea of managing incoming information by measuring performance on classic executive function tasks. Executive function is the ability to control one's attention on what they are doing by inhibiting distractors, focusing attention, and updating information in working memory (Banich, 2009; Diamond, 2012). Classic tasks of executive function, like Stroop, go/no-go, n-back, and flanker tasks, all present relevant and irrelevant information to participants and ask them to report on only the relevant parts of the information, ignoring any conflicting irrelevant information. As such, these types of tasks determine whether an individual can manage incoming information by sorting between relevant and irrelevant information and responding accordingly.

The present studies also operationalized the idea of the ability to hold incoming information by measuring performance on classic working memory capacity tasks. Working memory capacity is the amount of information an individual can store and process at one time (Baddeley & Hitch, 1994; Diamond, 2012). Classic tasks of working memory capacity, such as digit span, Corsi block, and operation span tasks, present sequentially increasing amounts of information for participants to hold and report back. As such, these types of tasks determine how much information an individual can hold and process at one time by identifying the point at which a learner can no longer hold a sequence of information correctly.

Theories of Working Memory

Working memory is a component of the system of memory that information must move through in order to be encoded into long-term memory and available for use in future situations (Atkinson & Shiffrin, 1968). However, researchers have different conceptualizations of what working memory is, ranging from a system with different components that hold and coordinate information (Baddeley & Hitch, 1974) to a system that retains information for a short period of time through attention (Engle et al., 1999; Kane et al., 2001) to a set of activated memory elements (Cowan 1999; Cowan et al., 2005). One thing that is consistent across these theories and research on working memory is that working memory is highly limited (e.g., Luck & Vogel, 1997; Miller, 1956). This section briefly explores these three main theories of working memory and their distinctions, although the goal is not to select which among these is the best theory.

One of the main theories discussing working memory was proposed by Baddeley and Hitch (1974). This model, also known as the multicomponent model of working memory, emphasizes how there are different subsystems that operate to hold information in working memory, such as the visuospatial sketchpad and the phonological loop. The central executive controls these two subsystems and directs how information flows into them. Additionally, the episodic buffer was added to this model later, and its function is to integrate spatial and temporal information with the other subsystems of working memory

(Baddeley, 2000). An emphasis of this theory is that all of the systems in the working memory model are severely limited in capacity and can hold only a small amount of information at a time.

In a second main theory, Engle et al. (1999) proposed a different theory of working memory focused on the control of attention, called the theory of controlled attention. This theory emphasizes the role of attention and attentional control in the processing of information. This theory explains how working memory operates through attentional processing, in which something enters working memory when it is attended to and leaves working memory when it is not attended to. In this theory, attentional control is limited in the amount of information that can be attended to at one time, and thus only so much information can be processed at any given time.

Third, Cowan (1999; Cowan et al., 2005) proposed a different theory, called the embedded processes model of working memory. In this theory, working memory is made up of what information is activated and attended to within long-term memory at any given time. However, this activation is limited in how long it can be activated for and limited in the amount of information that it can focus on at one time. Thus, only so much information can be activated in working memory at one time.

Across these main theories, other similar theories, and research on working memory, researchers agree that working memory, in whatever conceptualization they individually use, is limited in some way. Because of this limitation, there are two approaches to how working memory can operate to support learning when there is extraneous information present: (1) be able to determine which information is relevant and which is not and ignore the irrelevant information (i.e., better executive function) or (2) be able to process a large amount of information (i.e., larger working memory capacity). The present set of studies focuses on these two approaches to support learning despite the limits of working memory.

Components of and Measuring Working Memory

Executive Function: Managing Incoming Information. In this research, one's ability to manage incoming information is operationalized by measuring executive function. Executive function refers to cognitive processes, such as inhibition, updating, and task switching, that involve the use of cognitive and attentional control to pay attention to relevant novel stimuli (Banich, 2009; Diamond, 2012; Miller & Cohen, 2001). The ability to manage new incoming information through strong executive function is an important feature to consider in learning because if someone is able to sort out relevant from irrelevant information so they can focus solely on the relevant portion, they are more likely to learn better than someone who struggles to do that due to weaker executive function.

Additionally, individual differences in executive function may play an important role in multimedia learning. Although much of the research on variations in executive function have focused on disorders, like ADHD, autism, and depression (e.g., Barkley, 1997; Brown, 2009), there is also research that demonstrates that executive function varies across individuals within a neurotypical population as well (Friedman et al., 2008; Osaka et al., 2004; Parong et al., 2017). Research has suggested that individual differences in executive function can have an impact on learning, such that as executive function increases, the quality of learning outcomes increases as well (Albert et al., 2020; Grenell & Carlson, 2021).

This set of studies is concerned with individual differences in executive function as a way to operationalize how individuals differ in their ability to manage incoming information during multimedia learning. Thus, it is important to be able to measure these differences. There are many different ways to assess the strength of someone's executive function, and this set of studies uses brief versions of several classic measures, including the Stroop task, go/no-go task, n-back task, and flanker task. See Table 1 for a summary of the tasks used in these studies.

Stroop Task. The Stroop task is a measure of inhibition and attentional control in which participants are required to inhibit automatic processing of the meaning of a word and instead report on the ink color of a word (Diamond, 2012; MacLeod, 1991; MacLeod & MacDonald, 2000; Stroop, 1935). In the Stroop task, participants are presented with one of four color words: red, yellow, green, or blue. These words can be written either in the same color ink–for example, the word red written in red ink, which would be considered congruent–or can be written in a different color ink–for example, the word red written in yellow ink, which would be considered incongruent. The participant's task is to indicate the color of the ink. When the trial is congruent, this is generally an easy task because the participant does not have to sort out and ignore any irrelevant information or inhibit an automatic response. However, when the trial is incongruent, participants must inhibit the automatic response to indicate the meaning of the word and instead focus on the ink color. In incongruent cases, they must sort the relevant information–the color of the ink–from the irrelevant information–the word–and respond only to the relevant information. Thus, people who are better at managing the relevant information from the irrelevant information should

Table 1

Summary of Each Working Memory Task Used Across Experiments

Name	Description	Score
Stroop Task	Measure of executive function (inhibition); participants report the color of the ink a word is written in, which is either congruent with the meaning of the word or incongruent.	Average reaction time for incongruent trials minus average reaction time for congruent trials
Go/No-Go Task	Measure of executive function (inhibition); participants respond when stimulus says "go" and do not respond when stimulus says "no go".	Total number of errors on no-go trials
N-Back Task	Measure of executive function (inhibition and updating); participants see letters flashing and are told to respond with 'm' when the letter presented currently is the same as the letter presented 3 letters ago, otherwise response with 'n'.	Hits minus false alarms
Flanker Task	Measure of executive function (inhibition and attentional control); participants respond according to a stimulus presented in the middle of the stimuli, surround information is either congruent or incongruent.	Average reaction time for incongruent trials minus average reaction time for congruent trials
Forward Digit Span	Measure of working memory capacity; participants remember and report lists of numbers presented in specific order.	Total number of correct sequences divided by total number of sequences presented
Corsi Block Task	Measure of working memory capacity; participants remember and report sequences of blocks lighting up in specific order.	Highest number of blocks successfully remembered
Operation Span Task	Measure of working memory capacity; participants remember and report a sequence of letters presented in between math equations the participant determines are either correct or incorrect.	Total number of correct letters in correctly presented sequences

be able to respond to incongruent trials more quickly and more accurately than those who are worse at managing information.

Go/No-Go Task. The go/no-go task is also a measure of inhibition, specifically assessing withholding or stopping a prepared response (Diamond, 2012; Wright et al., 2014). For this task, participants are presented with two different signals, the go signal that tells the participant to respond as quickly as possible and the no-go signal that tells the participant to not respond at all. The go signal is presented about four times more frequently than the no-go signal, creating a situation in which participants need to respond repeatedly in quick succession until a no-go signal arises. This is a measure of inhibition because participants are given the go signal much more frequently and told to respond as quickly as possible, which makes it difficult to stop the automatic response of continuously responding when faced with the no-go signal. As such, people with better executive function will make fewer errors in not responding to the no-go signals.

N-back Task. The n-back task requires several different aspects of executive function to successfully complete the task. In the n-back task, a series of stimuli, usually letters, are presented one at a time (Diamond, 2012; Friedman et al., 2008; Kwong See & Ryan, 1995; Miller et al., 2009; Waris et al. 2017). Participants are tasked with reporting whether the stimuli currently presented matches the stimuli that was presented n stimuli ago. For example, in a 3-back task, participants are told to report whether the current stimulus matches the stimulus that was presented three stimuli ago. The n-back task has been used as a measure of updating ability (Waris et al., 2017) as participants are required to continually update the sequence of stimuli they are holding on to, which requires participants to manage which of the stimuli is still relevant and which of the stimuli is now irrelevant. Additionally, it can measure inhibition and attentional control as participants are required to prevent themselves from responding to incorrect responses (Diamond, 2012; Friedman et al., 2006; Miller et al., 2009). As such, a participant who is better at managing information in their working memory should be able to remember the relevant letters, like those presented in the last three trials, and ignore or forget the irrelevant letters, like those presented more or less than three letters ago, to perform better on this task than someone who is worse at managing incoming information.

Flanker Task. The flanker task is a measure of inhibition and attentional control (Diamond, 2012; Eriksen & Eriksen, 1974). This task can be done in several different ways, but generally, there is a target in the middle of a line of stimuli. This target indicates to the participant to make one of two different responses, depending on what it looks like. However, the target is surrounded by other information that either matches the target, in a congruent trial, or does not match the target, in an incongruent trial. Similar to the Stroop task, congruent trials are easier because all the information present on screen matches the information being portrayed by the target. In the incongruent trial, there is conflicting information because the target does not match the rest of the information on the screen. As such, those who are better able at managing information by sorting relevant from irrelevant information should be able to respond more quickly during incongruent trials than those who are worse at managing information.

Working Memory Capacity: Holding Incoming Information. For this set of studies, someone's capacity for holding incoming information is operationalized by working memory capacity. Working memory capacity is the amount of information that one can hold in their active consciousness and work with at any one time (Baddeley & Hitch, 1994; Diamond, 2012). The ability to hold incoming information through a large working memory capacity is an important feature to consider in learning because if someone is able to hold and process more information in working memory at one time due to a larger capacity, they would be better able to learn more information than someone who can hold and process less information at one time.

The limits of working memory capacity are not uniform across all individuals; some people are able to process and hold more information than others (Barrett et al., 2004; Ilkowska & Engle, 2010). Individual differences in working memory capacity can influence how much new information someone can process at one time and have an impact on how much of that information can be stored for later use. Because this set of studies examines individual differences in working memory capacity as a way to operationalize how individuals differ in the capacity they have to hold and process new information, it is important to be able to measure these differences. As with executive function, there are many different ways to assess someone's working memory capacity, and this set of studies uses several classic measures, including digit span, Corsi block task, and operation span task. See Table 1 for a summary of all tasks used in these studies.

Digit Span Task. The digit span task, also referred to in other literature as the forward digit span task, is a type of simple span task (Conway et al., 2005; Waris et al., 2017). Simple span tasks are thought to measure working memory storage capacity. Specifically in the digit span task, participants are shown a series of numbers, one number at a time, and participants are asked to report the exact sequence in the order the numbers were presented. On subsequent trials, the sequence becomes increasingly longer, one digit at a time. Participants are asked to do this task until they can no longer report a sequence correctly. This task helps determine how much information someone can hold in their working memory at one time without confusing or rearranging the numbers. At the point at which a number sequence can no longer be correctly reported, it can be deduced that the number of digits in the last correctly reported sequences is the holding capacity of a participant. As such, someone with a larger holding capacity should be able to successfully remember and report longer sequences than those who have a smaller holding capacity.

Corsi Block Task. The Corsi block task is a measure of working memory capacity (Corsi, 1972) and is specifically a measure of visuospatial working memory capacity. Participants are shown a display with nine blocks presented at different locations on the screen, then the blocks light up one at a time to create a sequence. On subsequent trials, the sequence becomes increasingly longer, one block at a time. Participants do this until they can no longer report the sequence correctly. This task helps determine how much information someone can hold in their working memory, specifically their visual working memory. At the point at which the visual sequence can no longer be correctly reported, the length of the last successfully reported sequence is the learner's working memory capacity. As such, someone with a larger holding capacity should be able to remember a longer sequence than those who have a smaller holding capacity.

Operation Span Task. The operation span task is a type of complex span task. Complex span tasks are similar to simple span tasks in that they both tax the holding capacity of working memory, but the complex span task also taxes processing capacity (Daneman & Carpenter, 1980; Ilkowska & Engle, 2010). This type of task may be a richer measure of working memory capacity when compared to simple span tasks because it requires processing in addition to storing information. This is more similar to how we normally use working memory capacity in our daily lives (Waris et al., 2017), and thus it may be a better indicator of differences in higher cognitive function (Engle et al., 1999). Specifically, for the operation span task, math equations are presented to learners and between these equations, single letters are presented and must be remembered in the sequence presented. The participant must hold the sequence of letters in mind while also solving math equations until prompted to report the sequence of letters. This task helps determine how much information someone can hold, i.e., the sequence of letters, while also being asked to work with other information, i.e., solve equations. The number of letters from correctly reported sequences can be used to determine someone's working memory capacity as those who have bigger working memory capacities should be able to report more sequences correctly as well as more accurately report longer sequences. The more successful the participant is at reporting the sequences, the bigger their working memory capacity is determined to be.

Role of Working Memory in Technology-Based Learning

Because working memory is highly limited in the amount of information an individual can take in at one time, it is essential to think about how material is presented to learners and what sorts of demands an instructional lesson can put on learners. This is especially true to consider when utilizing different types of technologies for instruction because new technologies can add negative demands on learners' working memories (Fenesi et al., 2015). For example, multimedia learning lessons, especially those that are poorly designed, can present information that is irrelevant to the goal of the lesson, which can pull attention away from the important parts of the lesson and hurt learning (Ayers & Paas, 2007). Another issue with some learning technologies is that the information they provide is often transient in nature. This means that the information is moving at a fast pace and continuously changing, which requires learners to hold onto certain parts of the material while also trying to process new incoming information (Kalyuga & Liu, 2015). If learners are using their limited working memory processing power to engage with information that is a distraction from the learning goal of the lesson, it means that they will not have as much available capacity left in working memory to engage with the relevant and important parts of the lesson. When this occurs, it is called *extraneous processing.* This makes meaningful learning less likely to occur.

However, when considering the limits of working memory, technology-based learning can be enhanced by following multimedia design principles (Mayer, 2021, 2022). When in use, several principles–such as the coherence, signaling, and redundancy principles –help reduce extraneous processing when using multimedia lessons and improve learning outcomes. The coherence principle proposes that individuals learn best when information that is not relevant to a lesson is removed from the instructional material. The signaling principle proposes that individuals learn best when the most important information in a lesson is pointed out, such as by arrows, color, highlighting, etc. The redundancy principle proposes that individuals learn best when verbal information presented through audio is not also shown visually as text on screen, as this adds to the load on working memory. These principles can be used in various ways to enhance how learners interact with instructional material in order to better learning outcomes.

Immersive Virtual Reality and Learning

The main technology of interest for this set of experiments is immersive virtual reality. Immersive virtual reality (IVR) is a technology that uses a head-mounted display and accompanying software to present images to the learner that help simulate being in a completely different environment than the one the learners are physically in. As learners move around and turn their heads, the head-mounted display tracks their movement and changes the scenery in front of their eyes to simulate being fully immersed in another environment. As the Cognitive Affective Model of Immersive Learning explains, using IVR for learning benefits students' experience of presence and agency due to the high control allowed for learners in this type of environment (Makransky & Petersen, 2021). As such, having an increased experience of presence and agency, learners should have increases in interest, intrinsic motivation, self-efficacy, embodiment, and self-regulation, all of which can have benefits for learning outcomes. Many studies have supported this by demonstrating that learning from IVR lessons can increase presence, interest, and/or motivation in learning (e.g., Lee et al., 2017; Makransky et al., 2019a; Parong & Mayer, 2018, 2021a, 2021b; Stephan et al., 2017). However, high presence and agency can also increase the cognitive load a learner experiences, specifically an increase in extraneous processing, which can hurt learning outcomes (Makransky & Petersen, 2021; Mayer et al., 2022).

A reason that IVR can increase cognitive load is through the increase in distracting visual features presented in an IVR lesson. When a learner is in a more traditional learning environment, like in a classroom or watching an instructional video, the learner knows that the information that they should focus on is right in front of them, on a screen at the front of the class or on their computer. However, this is not true for an IVR lesson; in IVR, visual information is presented 360 degrees around the learner. Thus, the information that was presented in a smaller, predictable area in a slideshow or in class is now being presented across 360 degrees of space in IVR. With all this visual input spread across more space, learners have to navigate around the environment to find the important information. Furthermore, with an increased amount of area where visual information can be present, an increased amount of visual information needs to be added into the IVR lesson, which means an increase in information that is likely not relevant to the goal of the lesson, i.e., distracting material.

As learning from IVR can pose a threat to limited working memory, it is important to study and develop a better understanding of the merits and pitfalls of using this technology for learning. A portion of the research investigating the impact of IVR on learning has found that using IVR can increase cognitive load on learning (e.g., Huang et al., 2020; Makransky et al., 2019b; Parong & Mayer, 2018, 2021a). As an example, learners wore EEG sensors during a lesson either while learning in IVR or from a 2D computer simulation; then, they switched learning mediums during the second part of learning. In the second part of the lesson, those in the IVR condition had higher load as indicated by EEG measurements, compared to the computer simulation condition (Makransky et al., 2019b).

Additionally, in another set of studies, learners saw a lesson on the bloodstream either in IVR or a slideshow (Parong & Mayer, 2021a). When learners saw the IVR lesson, the EEG measures indicated that those in the IVR lesson were in a lower cognitive engagement state compared to the slideshow condition and participants reported higher subjective cognitive load in the IVR lesson compared to the slideshow lesson. Furthermore, a review on literature investigating the role that cognitive load plays in learning reported that those who learned with IVR have increased cognitive load (Han et al., 2021).

As IVR increases cognitive load, this can hurt learning which has been demonstrated in a portion of the literature comparing learning via IVR to learning from more traditional lessons. For example, students were taught about crime scene investigation concepts either in IVR or using a video lesson (Makransky et al., 2020). Those who learned the material using IVR performed significantly worse on a declarative knowledge posttest than those who learned with the video lesson. In another study, participants learned about the blood stream either in IVR or with a slideshow lesson (Parong & Mayer, 2018, 2021a). Those who learned with the IVR lesson did worse on a posttest compared to those who learned using a slideshow lesson. This was also demonstrated in a similar study, but with a history lesson in IVR compared to a video lesson (Parong & Mayer, 2021b).

Not all studies have demonstrated that IVR is hurtful for learning. Some studies have demonstrated that IVR can enhance learning compared to a more traditional lesson (e.g., Alhalabi, 2016; Calvert & Abadia, 2020; Makransky et al, 2019b; Webster, 2016). Additionally, some studies have shown that IVR is equivalent to more traditional lessons in terms of learning outcomes (e.g., Ekstrand et al., 2018; Kozhevnikov et al., 2013; Lee et al., 2017; Makransky et al., 2020). However, when the literature is analyzed through meta-analyses, studies show a generally small but positive effect size of using IVR for learning (e.g., Coban et al., 2022; Wu et al., 2020).

The conflicting findings from different literatures make it difficult to interpret which aspects of this type of technology are beneficial for learning and which parts are not. There are several reasons that may explain these conflicting findings, including the types of lessons presented and the types of resources learners have while learning. Lessons presented across different experiments are not always equivalent in their quality (Flavia di Natale et al., 2020; Hamilton et al., 2021). Some experiments may be presenting learners with well-designed and low distracting lessons while other experiments may be presenting learners with distracting and poorly-designed lessons. As such, when trying to draw conclusions about the benefits of lessons in IVR, these differences likely impact the benefits on learning outcomes and create a disconnect in the literature. Additionally, learners have different types of mental resources to learn in these types of environments, like differences in prior knowledge, cognitive systems, etc., that may play a role in learning from IVR lessons. Thus, without further research that investigates and/or controls for these factors, it is difficult to have any conclusive decisions about the use of IVR in learning. The present set of experiments aims to understand how this second point may be impacting findings in IVR research.

Mitigating Distractions in Learning

Coherence Principle. One possible reason for what could be hurting learning and/or adding cognitive load while using IVR lessons is that learners may experience extraneous processing while engaging with the lesson. In an IVR lesson, a learner is surrounded by 360-degrees of perceptually engaging information, much of which is attention-grabbing but irrelevant. The information that is relevant to the goal of the lesson makes up only a portion of everything the learner can be attending to, which may make it more difficult for participants to determine and separate what is important from what is less important for understanding. Because of this, it is likely that some of the information that they are processing is irrelevant, which increases extraneous processing and lowers the amount of capacity that can be dedicated to processing relevant information within working memory.

Several principles have addressed how to reduce extraneous processing, but the one that this research focuses on is the coherence principle (Fiorella & Mayer, 2022; Mayer, 2021). The coherence principle explains that information that is unrelated to the instructional goal of a lesson should not be included in the lesson, as this adds to the amount of information that needs to be processed, thereby creating extraneous processing. Instead, only the information that is relevant to the lesson should be presented to learners in order to allow for the best chance at learning to occur.

Prior research has demonstrated the benefits of using the coherence principle to benefit learning. For example, Harp and Mayer (1997) gave students a lesson that was made more interesting by including entertaining information–interesting textual information, interesting pictures, or both–or a base lesson that only included the relevant text and graphics. Those who saw the base version of the text remembered significantly more information from the text than all other groups. Similarly, in a study that investigated if adding sounds and background music to a lesson on the formation of lightning impacted learning, students who learned from lessons with background music and sound effects, just background music, and just sound effects all performed worse on retention and transfer tests compared to students who had neither background music nor sound effects (Moreno & Mayer, 2000). In reviews of literature investigating this effect, large effect sizes have been found in support for the coherence principle (i.e., Fiorella & Mayer, 2022; Mayer, 2021; Rey, 2012), suggesting that getting rid of irrelevant information has a strong benefit for learning.

The coherence principle can play a role in understanding the potential difficulties of learning in IVR, as much of the visual and auditory information in IVR lessons is irrelevant to the learning goals. As such, many lessons in IVR likely violate the coherence principle. One way to combat this is to not present the information in IVR and instead present it using a less distracting medium, such as a slideshow.

Pre-training. Another possibility for why it may be more difficult to learn from IVR lessons is because IVR is a new technology for many people, and a new medium to learn from for most learners. This means that learning in an IVR lesson requires participants to not only learn the information being presented in the lesson, but also learn how to use the VR system itself, such as moving around, using the handheld controls, etc. Thus, learning how to use the VR system can act as a source of extraneous processing because it is information that the learner can be distracted by that is not relevant to the lesson itself.

There are several principles that can be followed to deal with this type of issue, but the focus of this research is on the pre-training principle (Mayer, 2021; Mayer & Fiorella, 2022). The pre-training principle explains that certain information, like keywords, can be introduced prior to a lesson in order to reduce the amount of new information presented during the lesson. This moves the processing of a portion of the information, e.g., keywords, to before the lesson and allows the learners to focus on the interrelations between different parts of the material during the actual lesson.

Prior research has shown benefits for using pre-training in learning environments. For example, in a course on electrical engineering, students either learned first about how each component of an electrical appliance worked then learned how all the pieces worked together or spent the whole time learning about how all the components worked together (Pollock et al., 2002). Those with low prior knowledge who were taught following the pre-training principle (the first group) performed significantly better on a transfer test than those who did not follow the pre-training principle. Similarly, another study taught participants about how a car's braking system worked (Mayer et al., 2002; Exp. 1). Participants in the pre-training group learned first about each part of the brake system then learned about how the whole system works together. Those in the no pre-training condition learned about the information altogether. Those in the pre-training condition performed significantly better on the posttest than those who were in the no pre-training condition. Research investigating the pre-training principle across literature has found that pre-training has a consistent, positive effect on learning (Mayer & Fiorella, 2016, 2022).

Typically, the research on pre-training has investigated how presenting some of the lesson content prior to the lesson can benefit learners. However, for learning in IVR, it may be more beneficial to give the learners an opportunity to understand how the technology works prior to starting a lesson in IVR. This should benefit learners, especially those new to

VR technology, if learning how to use the VR device creates extraneous processing on the learner. If the learner can engage with the technology prior to the lesson and use their cognitive capacity during the lesson to focus on the material, they should be able to learn the material better. Although slightly different from how the pre-training principle has been investigated in the past, this is an extension of the pre-training principle. In this version of pre-training, information that does not need to be presented during the lesson, i.e., information on how to use the VR device, can be off-loaded and presented prior to the lesson, which leaves only the relevant content-specific information for the learner to focus on during the lesson.

Theoretical Background

Cognitive load theory (CLT; Paas & Sweller, 2022; Sweller, 1994; 2020) presents a way to understand how different parts of instructional materials impose load in different ways on limited working memory in order to understand how to create better instructional material. Similarly, the cognitive theory of multimedia learning (CTML; Mayer, 2021; Mayer & Fiorella, 2022) presents a similar perspective on the use of cognitive capacity in multimedia learning environments. Because there is a limited amount of cognitive capacity for individuals to use to process new information, due to the limitations of working memory, the way in which this capacity is allocated is vital. Both theories present three ways in which information can add load to learners, each having a different amount of desirability in learning.

The first type of processing is called extraneous load (as per CLT) or extraneous processing (as per CTML) which occurs when learners process information that is not relevant to the goal of the lesson (Mayer, 2021; Mayer & Fiorella; 2022; Paas & Sweller, 2022; Sweller, 1994, 2020). This occurs when there are irrelevant words or images in the lesson. Additionally, this can come from a learner trying to understand how to use a new technology for instruction. This type of processing is not desirable in learning as processing information that is irrelevant to the lesson itself takes away processing that should be directed towards important parts of the lesson. As such, to create a learning environment that is beneficial for learning, extraneous processing should be minimized as much as possible.

The second type of processing is called intrinsic load (CLT) or essential processing (CTML) which occurs when learners are working to create a coherent mental model of the relevant information in their working memory (Mayer, 2021; Mayer & Fiorella; 2022; Paas & Sweller, 2022; Sweller, 1994, 2020). This kind of processing occurs with all lessons but increases as material becomes more complex and/or the components of the lesson become more interrelated. Although this type of processing is required for learning to occur, the goal is to manage the amount of essential processing in order to allow enough space in working memory for generative processing, a more desirable type of processing, to occur.

The last type of processing is called germane load (CLT) or generative processing (CTML) which occurs when learners try to make sense of the material and make connections within the material and between the material and ones' prior knowledge (Mayer, 2021; Mayer & Fiorella, 2022; Paas & Sweller, 2022; Sweller, 1994, 2020). This type of processing occurs when a learner actively draws connections between different parts of the material or makes connections between their prior understanding and the new material. Generative processing is the most desirable type of processing because it leads to a deeper and longer-lasting understanding of the material. However, it does not necessarily occur during the learning process, so the goal with this type of processing is to encourage learners to engage in it while learning.

The aim of this set of experiments is to extend cognitive load theory and the cognitive theory of multimedia learning to incorporate how individual differences in both the ability to hold and manage incoming information impact learning. By incorporating individual differences related to working memory into cognitive load theory and the cognitive theory of multimedia learning, these theories could make better predictions about learning performance in various environments.

Chapter II: Experiment 1

Rationale and Hypotheses

The goal of Experiment 1 was to understand how individual differences in how well a learner manages incoming information (executive function) and how much information the learner can hold at one time (working memory capacity) play a role in learning from online multimedia lessons that vary in their level of distractibility. The main question being addressed with this research was: do executive function and working memory capacity predict learning across learning environments with various levels of distraction? I investigated this question in a series of three correlational studies in which participants had to learn from a video lesson with low, moderate, or high levels of distraction, and then completed a posttest and tasks to measure individual differences in executive function and working memory capacity.

The first hypothesis predicted that learners who are better at managing incoming information and are able to hold more information at one time should develop a better understanding of the material. This is because those who are better able to manage and hold information will be better at ignoring irrelevant information and have more cognitive resources for processing the relevant information. More specifically, it is predicted that posttest scores will correlate with scores on executive function measures (hypothesis 1a) and working memory capacity measures (hypothesis 1b). This hypothesis was tested in Study 1a and Study 1b. However, this should not be true when a lesson does not impose demands on these characteristics. As such, it is also predicted that at low levels of distraction, there should not be a correlation between posttest scores and scores on executive function (hypothesis 2a) and working memory capacity measures (hypothesis 2b). This hypothesis was tested in Study 1c. This pattern of results is predicted because if a lesson has fewer distractions (i.e., it does not increase extraneous processing by presenting information that is irrelevant to the goal of the lesson), learners' ability to manage incoming information and/or their ability to hold and process more information should not matter as much as when there are more opportunities to engage with extraneous processing (i.e., in lessons that have distracting features). Across the three studies, this should manifest itself in the correlations for working memory capacity and executive function with posttest performance to be significantly stronger in lessons with distractions than in lessons without distractions (hypothesis 3).

In comparing group means across the three studies, it is predicted that learners will report experiencing higher extraneous processing with lessons that have more distracting features than those with fewer distracting features (hypothesis 4). Lastly, lessons that have more distracting features should also result in lower scores on tests of learning than lessons that have fewer distracting features (hypothesis 5).

Study 1a (High Distraction Lesson)

The first study focused on understanding how learners' ability to manage and hold incoming information, measured specifically by classic tasks of executive function and working memory capacity, related to learning from a lesson with a high level of distraction. Specifically, I examined the correlation between scores on executive function and working memory capacity tasks and posttest scores after learning from a lesson with many distracting features.

Method

Participants and Design. There were 152 participants recruited from the Psychology Subject Pool at the University of California, Santa Barbara. The mean age of the participants was 18.90 years (*SD* = 3.23) and most of the participants were in their 1st or 2nd year at the university (*M* = 1.99 years, *SD* = 1.00). Additionally, the prior knowledge of the participants, on a 5-point Likert scale from “1 – Very Low” to “5 – Very High,” was somewhat low (*M* = 1.70, *SD* = .99). Of the sample, 110 participants identified as women, 40 participants identified as men, one participant identified as non-binary, and one participant did not indicate their gender. The predictor variables for this study were executive function task scores, including scores on the Stroop task, go/no-go task, and n-back task, and working memory capacity task scores, including scores on the digit span task and Corsi block task. The outcome variable for this study was posttest score.

Materials. All of the materials were computer-based and presented either on Qualtrics or on PsyToolKit (Stoet, 2010, 2017). These materials included a prequestionnaire, lesson, filler task, posttest, postquestionnaire, and working memory tasks.

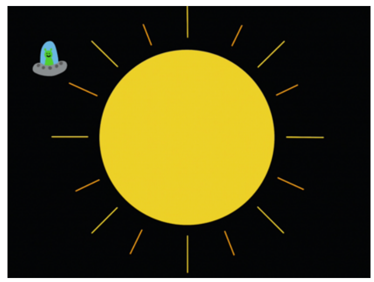
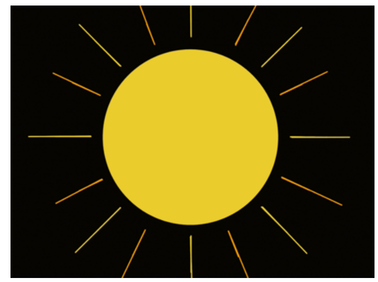
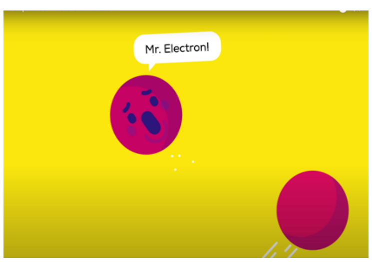
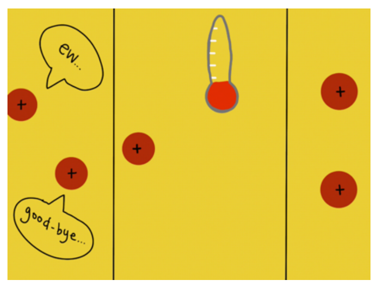
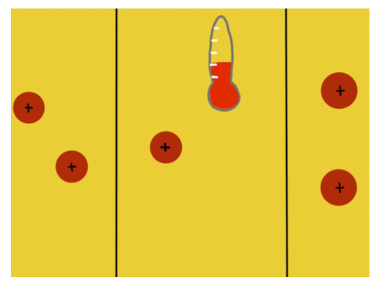
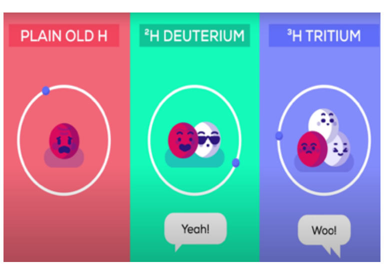
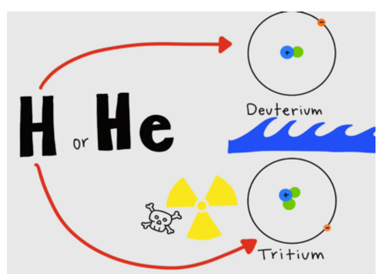
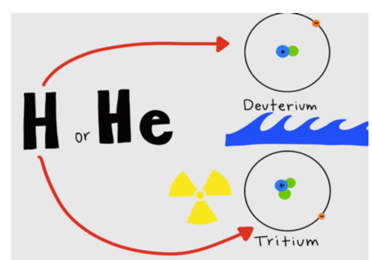
Prequestionnaire. The prequestionnaire collected information about the participants, including their age, gender, class level, and prior knowledge. Prior knowledge on fusion energy was assessed on a 5-point Likert scale from “1 – Very Low” to “5 – Very High.” This was used to gather information about each participant’s subjective prior knowledge of the topic instead of using a pretest in order to not introduce testing effects into the study and give students information about what knowledge to focus on during the lesson (Karpicke & Aue, 2015; McDaniel et al., 2007; Rowland, 2014). There was also a question that asked participants to check if they had completed certain activities that could more objectively determine someone's prior experience with the topic of the lesson, but this was not used due to the extremely low reliability; McDonald's Omega = .23 (Hayes & Coutts, 2020). The prequestionnaire is presented in Appendix A.

Lesson. Participants watched a lesson on fusion energy from YouTube, found here: https://www.youtube.com/watch?v=mZsaaturR6E&ab_channel=Kurzgesagt%E2%80%93InaNutshell. In the lesson, participants watched an animation with a voice over that discussed how fusion energy works in the sun, how fusion energy may be adapted to be useable on Earth, and the pros and cons of using fusion energy as an alternative to other forms of energy. This lesson was selected for use in this study because it included many seductive details, presented both visually and auditorily. For example, visually, there was a lot of irrelevant movement of the images, vibrant and unnecessary colors, and animations that did not relate to the main goal of the lesson. Verbally, there were many interesting but irrelevant statements presented to participants in the voice over and there were unrelated background sounds. Additionally, the lesson moved quickly through the different animations. The amount of verbal and visual seductive details and the transience of the presented information is why this lesson was considered distracting; these details likely could made it more difficult to focus on the information that was relevant for the goal of the lesson. The lesson lasted a total of 6 minutes. Images from the lesson can be seen in Figure 1.

Filler Task. The filler task was used to delay the time between watching the lesson and taking a test on the material. It was also used as a way to stop participants from keeping information in working memory through rehearsal and allowed the posttest to assess

Figure 1.

Images from the Various Lessons in Experiment 1

High Distraction (Study 1a)	Moderate Distraction (Study 1b)	Low Distraction (Study 1c)
		
		
		

information that had been fully encoded. For this task, participants were told they would be watching a video and answering questions about the presented video. The video showed a compilation of trick shots, such as throwing a carton of food over an aisle to land perfectly in a shopping card or throwing a snack bar into an unzipped backpack from across a room.

The video used for this can be found here:

https://www.youtube.com/watch?v=qlzVPauUgw8&ab_channel=DudePerfect. The participants could not move forward in the lesson before the video finished. Once the video finished, they could move to the next page that asked participants to describe their favorite trick shot from the video. This was used to (1) pressure participants to pay attention to this video instead of rehearsing previous information and (2) serve as a check that participants watched the video.

Posttest. The posttest consisted of eight open-ended questions to assess participants' understanding of the material presented in the lesson. All of the questions aimed to assess transfer of knowledge from the video at some level but varied in how much transfer they needed to engage in to successfully answer the question. The questions included (1) "How does fusion work in the sun?" (2) "Why is fusion energy so efficient?" (3) "Why can the sun create fusion energy without having a hot enough temperature?" (4) "What are the pros of using fusion energy on Earth?" (5) "What is one possible environmental reason to not use a fusion reactor?" (6) "What are the cons of using fusion energy?" (7) "What happens if the confinement of a fusion reactor fails?" (8) "Why does the reaction stop if the confinement of a fusion reactor fails?"

The posttest questions were presented one at a time for participants to respond to. They were given 90 seconds to respond to each question, which was generally more than sufficient time for participants to provide complete responses. They were not allowed to move forward before the 90 seconds had passed and once the time had passed, they were automatically moved to the next question. This time constraint was imposed for several reasons. First, since this study was done remotely, it helped ensure that participants could not skip through the questions without answering. Additionally, the time limit standardized the testing experience across all learners to determine which information was most accessible to learners at the time of answering. The posttest questions are listed in Appendix B.

Postquestionnaire. The postquestionnaire included 17 rating items. The first five questions assessed participants' experience with the lesson, including questions like "I enjoyed the lesson." and "I felt motivated to understand the material." The next nine questions assessed the participants' subjective cognitive load experience, specifically for extraneous, essential, and generative processing during the lesson (Parong & Mayer, 2018, 2021a, 2021b). These included questions like "I felt distracted during the lesson." for extraneous processing, "I was trying to learn the main facts from the lesson." for essential processing, and "I was trying to make connections between the material and things I already know." for generative processing. The McDonald's Omega was .84 for extraneous processing questions, .61 for essential processing questions, and .66 for generative processing questions. The postquestionnaire questions are listed in Appendix C.

Working Memory Tasks. A battery of working memory tasks was used to assess participants' ability to manage incoming information (executive function) and ability to hold new information (working memory capacity). Participants completed five classic working memory tasks to help assess individual differences in executive function and working memory capacity. The digit span task was presented on Qualtrics and each sequence length was completed twice. The other four tasks were presented on PsyToolKit (Stoet, 2010,

2017) and each task was completed twice. The first time the participants completed the task served as a training block then the second time they completed the task served as the real block. Only the real block responses counted for the participants' scores on each task.

Participants completed the *Stroop task*. On the screen, participants were told they would see the names of one of four colors (red, yellow, green, or blue) written in one of the four ink colors (red, yellow, green, or blue). They were told to report the ink color by using the keyboard to press "r" for red ink, "y" for yellow ink, "g" for green ink, or "b" for blue ink. Before the session started, they saw a fixation cross in the middle of the screen. About 1/3 of the trials were congruent trials (i.e., the printed name and ink color were the same) while the other 2/3 were incongruent (i.e., the printed name and ink color were different). Participants completed 40 trials in both the practice block and the real block. For this set of studies, the score of the Stroop task was determined by subtracting the average reaction time on the congruent trials from the average reaction time on the incongruent trials. Lower scores on this task indicated better executive function.

Participants completed the *go/no-go task*. On the screen, participants saw one of two images: either a go stimulus showing a green oval with the words "GO press the spacebar" or a no-go stimulus showing a red oval with the words "NO GO do not press the spacebar." The participants had to press the spacebar on the keyboard as quickly as possible when they saw the go stimulus and not respond at all when they saw the no-go stimulus. The cues were presented one at a time. Participants completed 25 trials in both the training block and the real block. For this set of studies, the score of the go/no-go task was determined by the number of errors made on the no-go stimuli. However, due to a ceiling effect with participants being too good at this task, scores were dichotomized such that they were labeled "0" if they made no errors on the task and labeled "1" if they made one or more errors on the task. Lower scores on this task indicated better executive function.

Participants completed the *n-back task*. On the screen, participants first saw a fixation cross then they were shown a series of letters. Participants were told their goal was to identify if the letter presented at that moment matched or did not match the letter presented three letters ago. They were told to press "m" if the letter did match and press "n" if the letter did not match. Participants completed 20 trials in both the training block and the real block, but only 17 of these were used for analyses as the first three letters were automatically non-matching letters. For this set of studies, the score of the n-back task was determined by subtracting the number of incorrect responses from the number of correct responses. Larger scores on this task indicated better executive function.

Participants completed the *digit span task.* On the screen, participants first saw a fixation cross. Then, a series of numbers presented at a rate of one number per second were presented. The sequence started with a length of three digits and increased to a length of 10 digits, with each sequence length being presented twice. After each sequence was displayed, participants were asked to report the sequence to the best of their ability in a textbox. Participants completed 16 trials, two trials for each sequence length. For this set of studies, the score for the digit span task was determined by the total number of correctly reported sequences divided by the total number of sequences presented, which was 16. Higher scores on this task indicated larger working memory capacity.

Participants completed the *Corsi block task.* On the screen, participants were shown an image of nine yellow blocks randomly arranged on the screen. Then, the blocks lit up purple one at a time to form a sequence. Once the sequence was complete, participants were told "Go" and instructed to click on the blocks in the same sequence previously displayed. The sequence started with three blocks and with every correct response, increased by one block up to a sequence of nine blocks. If the participant responded wrong to a sequence, they repeated that sequence length with a new sequence. If their response was correct on the second attempt at the sequence length, they would continue forward with the task, going onto a sequence with an additional block. However, if they were wrong again, the session was done. Participants completed a varied number of sequences in the training and real blocks, as the task would terminate once the participant made an error on two sequences of the same length. For this set of studies, the score for the Corsi block task was determined by the number of blocks in the longest sequence the participant could successfully repeat. Higher scores on this task indicated larger working memory capacity.

Apparatus. The first part of the study, including the prequestionnaire, lesson, filler task, posttest, postquestionnaire, and digit span task were all presented through Qualtrics. Qualtrics is a web-based survey platform that can be used to create and distribute online surveys. The remaining four working memory tasks were presented through PsyToolKit (Stoet, 2010, 2017). PsyToolKit is a free-to-use platform and toolkit for running cognitive experiments and surveys. Participants completed this study using their own laptops or desktop computers at home; participants were told they could not participate in this study if they were using a tablet or smartphone, due to limited features of the PsyToolKit tasks.

Procedure. On the day participants signed up to partake in the study, they were sent an email in the morning with a link to the first portion of the study. They were instructed to complete the study at some point during the day, all at one time, from their own laptop or computer. When they opened the link, they first completed the consent form. Then they were taken to a page with instructions on how to complete the entire experiment. Next, they completed the postquestionnaire and watched the video lesson. During the video lesson, participants were not able to move to the next page until they watched the full video. Then, participants completed the filler task, during which they could not move forward until they watched the entire video.

After the filler task, participants were taken to the posttest page in which they had 90 seconds to write an answer for each question. Once the 90 seconds were up, the next question was displayed. Once all eight questions were answered, the participants moved to the postquestionnaire, which was self-paced. Then, participants completed the digit span task. After the digit span task was complete, participants were given a second link to take them to the PsyToolKit page. On this page, they completed the Stroop task, go/no-go task, Corsi block task, then the n-back task. Each of these tasks was completed twice in a row. Finally, participants were thanked and given credit for participating. This study took no longer than an hour to complete and was approved by the university's Institutional Review Board (IRB).

Results

Posttest Scoring. The posttest was scored with a rubric that assessed how many main points were present in the responses the participants provided in the various responses to the posttest questions. For example, a question like “Why can the sun create fusion energy without having a high enough temperature?” had two main points associated with it and that could earn the participants points. Specifically, these main points were “high pressure is used instead of heat” and “pressure can squeeze together the nuclei.” As such, if participants did not earn all the points, this did not necessarily indicate that their answer was incorrect, but instead indicated how many of the relevant and correct details were included in their response to the question. Participants could respond to a question correctly (e.g., in the example question, a correct answer would say something like “because the sun has high pressure.”) but not include all the main points in the rubric for that item. This rubric style was used in order to determine who demonstrated more knowledge about the material by including more specific facts to support their response rather than a binary right/wrong scoring scheme. The main points associated with each question are provided in Appendix B.

Each question had two to seven main points associated with it. When a participant’s response included a main point from the rubric in their answer, the response was given a “1” for that specific rubric item, and if not, it was given a “0.” Then a total score for each question was created by adding the 1s and 0s across the rubric. Then, a total score for the posttest was created by adding the total scores for each question, to create a number out of 31. Participants were not expected to produce all the main ideas for each question.

Two independent scores read each response and gave the participants scores. The correlation between the two scorers was strong, $r = .90$, $p < .001$. All differences in scoring between the scorers was discussed until 100% agreement. The reliability of the posttest was determined using McDonald’s Omega and was .74. The lower Omega score is likely due to the fact that this test assessed knowledge requiring different levels of transfer, from near transfer to far transfer.

Outliers. In this data, outliers were defined as participant scores on the measures of working memory in which the z-score exceeded positive or negative three. In this data set, eight participants were removed from the data because of their outlier scores, including four participants due to their high go/no go task scores, three participants due to their low Corsi block task scores, and one participant due to their low n-back score. Removing these outliers did not change the pattern of results.

Hypothesis 1: Do Posttest Scores Correlate with Scores on Executive Function and Working Memory Capacity Tasks when Watching a Highly Distracting Lesson? Hypothesis 1 predicted that posttest scores should correlate with executive function and working memory capacity task scores as participants had watched a lesson that was high in distraction. The main analysis to test this hypothesis was to correlate posttest scores with scores on the executive function and working memory capacity tasks. A correlation matrix was constructed to determine the relationships between posttest scores and each of the five working memory tasks, which is displayed in Table 2. Scores on the n-back task were significantly correlated with posttest scores, $r = .31$, $p < .001$. See Figure 2 for scatterplot of n-back and posttest scores. None of the other executive function tasks were significantly correlated with posttest score. This is partially consistent with hypothesis 1a. None of the working memory capacity tasks were significantly correlated with posttest score, in contrast to hypothesis 1b.

Table 2

Correlation Matrix for Study 1a

	Stroop	Go/No-Go	N-Back	Digit Span	Corsi
Posttest	-.14	-.07	.31***	-.05	.03
Stroop		.11	.04	-.09	.10
Go/No-Go			-.03	.01	-.10
N-Back				.06	.16
Digit Span					.01
Corsi					

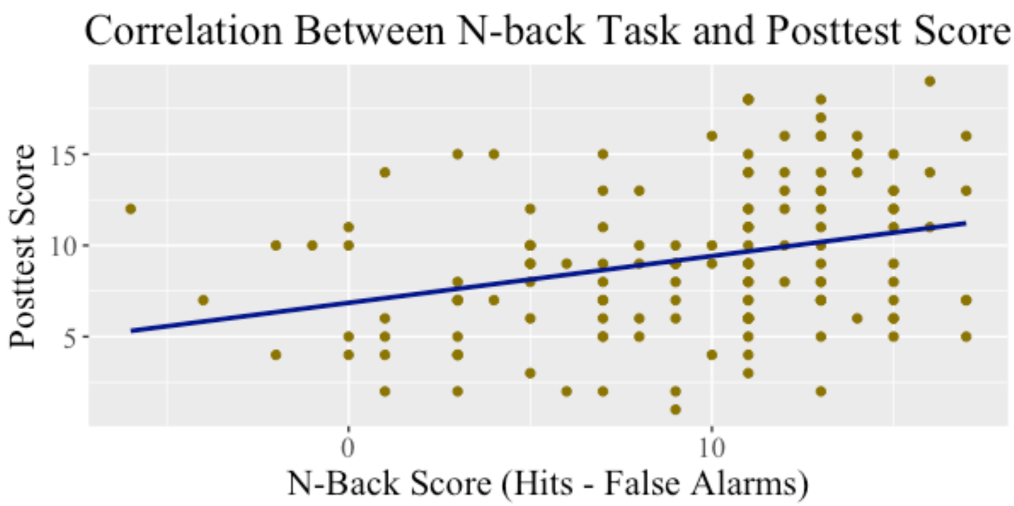
* $p < .05$, ** $p < .01$, *** $p < .001$
Lower scores on Stroop and go/no go tasks indicate better executive function

To supplement this analysis, a stepwise regression was used to determine if there was a predictive relationship between the scores on each of the five working memory tasks and posttest score for the highly distracting lesson. All five predictors were included in the regression and the program selected only those that provided predictive power in the final output. A stepwise regression was used to try to isolate only those tasks that were strongly correlated with posttest scores. Only the n-back task had predictive power for the posttest scores, so it was the only variable of the five that was included in the stepwise regression, $R = .31$, $R^2 = .09$, $F(1, 141) = 14.69$, $p < .001$. The n-back task was related to posttest performance in the predicted direction, such that as n-back score increased–indicating an increase in executive function–posttest scores also increased, $B = .26$, $\beta = .31$, $t = 3.83$, $p < .001$. Overall, there is some evidence that individual differences in executive function are related to learning outcomes from a highly distracting multimedia lesson.

Exploratory Analyses. Several exploratory analyses were conducted to understand how executive function and working memory capacity interact with some of the

Figure 2.

Correlation Between N-back and Posttest Scores for Study 1a



postquestionnaire responses participants made. See Table 3 for a correlation table. First, I examined the relationship between extraneous processing and performance on the five working memory tasks. The n-back task was significantly, negatively correlated with extraneous processing, such that as n-back score increased, meaning executive function increased, subjective experience of extraneous processing decreased. This suggests that having better executive function is related to learners being able to ignore distracting information thereby allowing them to focus only on the relevant information, as would be expected.

Next, I examined the relationship between essential processing and performance on the five working memory tasks. None of the tasks significantly correlated with essential processing scores. This suggests that essential processing may not relate to changes in managing or holding incoming information.

Table 3

Correlations between Cognitive Load and Working Memory Tasks in Study 1a

	Extraneous Processing	Essential Processing	Generative Processing
Stroop	.09	-.09	-.16*
Go/No-Go	.05	-.09	-.05
N-back	-.19*	-.10	.03
Digit Span	.10	-.08	-.07
Corsi	-.09	.04	-.02

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

Finally, I examined the relationship between generative processing and performance on the five working memory tasks. Stroop task score was significantly, negatively related to posttest performance, such that as Stroop tasks score decreased, meaning executive function increased, posttest score increased. This may suggest that better executive function is related to being better able to engage in generative processing. Someone who is better able to keep irrelevant information out of their working memory through better executive function should have more working memory space to engage in generative processing.

Study 1b (Moderate Distraction Lesson)

The second study focused on understanding how learners' ability to manage and hold incoming information, measured specifically by classic tasks measuring executive function distractions. Specifically, I examined the correlation between posttest scores and scores on executive function and working memory capacity tasks after learning from a lesson with some distracting features.

Method

Participants and Design. The participants included 74 participants recruited from the Psychology Subject Pool at the University of California, Santa Barbara. The mean age of the participants was 18.92 years (*SD* = 1.16) and most of the participants were in their 1st or 2nd year at the university (*M* = 1.73 years, *SD* = .90). Additionally, the prior knowledge of the participants was somewhat low (*M* = 1.74, *SD* = .86) based on a 5-point Likert scale from "1 – Very Low" to "5 – Very High,". Of the sample, 44 participants identified as women and 30 participants identified as men. The predictor and outcome variables for this study were the same as Study 1a.

Materials. All of the materials were the same as Study 1a, except for the lesson.

***Lesson*.** The lesson participants watched discussed the same material as the lesson presented in Study 1a, but the lesson was modified by the experimenter to lower the amount of distraction present. The lesson was still an animation with a voice over, but there was less visual information presented to the learner over the course of the lesson. This lesson reduced the transient property of the original video by having a looping animation in each scene of the lesson. In each scene, there were still several visually distracting features that the learners saw, such as a space creature or comet flying through the scene. The lesson was less distracting than the one in Study 1a due to the lowered amount of irrelevant visual information and the reduced pace of how the visual information was presented. That being said, there was still irrelevant information presented visually and the same verbal seductive details were present. The lesson lasted a total of 6 minutes. Images from the lesson can be seen in Figure 1.

Postquestionnaire. The postquestionnaire was the same as Study 1a. The McDonald's Omega was .86 for extraneous processing questions, .64 for essential processing questions, and .67 for generative processing questions.

Apparatus. The apparatus used were the same as in Study 1a.

Procedure. The procedure was the same as in Study 1a.

Results

Posttest Scoring. The posttest was scored in the same way that Study 1a was. The correlation between the two scorers was strong, $r = .90$, $p < .001$. McDonald's Omega for the posttest was .74.

Outliers. Outliers were dealt with in the same way as they were in Study 1b. Two participants were removed for having extreme scores; one participant for having a high score on the Stroop task and one participant for having a high score on the go/no-go task. Removing these outliers did not change the pattern of results.

Hypothesis 1: Do Posttest Scores Correlate with Scores on Executive Function and Working Memory Capacity Measures when Watching a Moderately Distracting Lesson? Hypothesis 1 predicted that posttest scores should correlate with executive function and working memory capacity task scores as participants had watched a lesson that included a moderate amount of distraction. The main analysis to test this hypothesis was to correlate posttest scores with scores on the executive function and working memory capacity tasks. A correlation matrix was constructed to determine the relationship between posttest scores and each of the five working memory tasks, which is displayed in Table 4. As in Study 1a,

Table 4

Correlation Matrix for Study 1b

	Posttest	Stroop	Go/No-Go	N-Back	Digit Span	Corsi
Posttest	1.00	-.18	-.22	.26*	-.09	-.05
Stroop		1.00	.18	-.26*	-.12	.06
Go/No-Go			1.00	-.23*	-.19	.03
N-back				1.00	..22	.15
Digit Span					1.00	.09
Corsi						1.00

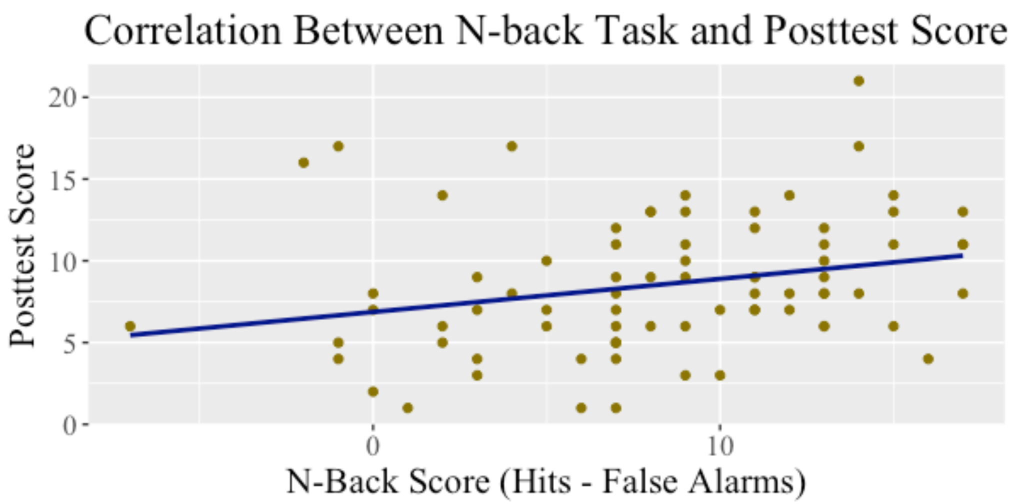
* $p < .05$, ** $p < .01$, *** $p < .001$
Lower scores on Stroop and go/no go tasks indicate better executive function skill

scores on the n-back task were significantly correlated with posttest scores, $r = .26$, $p = .027$. See Figure 3 for scatterplot of n-back and posttest scores. None of the other executive function tasks were significantly correlated with posttest score, partially consistent with hypothesis 1a. None of the working memory capacity tasks were significantly correlated with posttest score, in contrast to hypothesis 1b.

To supplement this analysis, a stepwise regression was used to determine if there was a predictive relationship between the scores on the five working memory tasks and posttest score for the moderately distracting lesson. All five predictors were included in the regression and the program selected only those that provided predictive power in the final output. Once again, only the n-back task had predictive power for the posttest scores, so it was the only variable of the five that was included in the stepwise regression, $R = .26$, $R^2 = .07$, $F(1, 73) = 5.09$, $p = .027$. The n-back task was related to posttest performance in the predicted direction, such that as n-back score increased–indicating an increase in executive function–posttest scores also increased, $B = .20$, $\beta = .26$, $t = 2.26$, $p = .027$. In line with

Figure 3.

Correlation Between N-back and Posttest Scores for Study 1b



Study 1a, there is some evidence that individual differences in executive function are related to learning outcomes from a moderately distracting multimedia lesson.

Exploratory Analyses. Several exploratory analyses were conducted to understand how executive function and working memory capacity interacted with some of the postquestionnaire responses participants made. See Table 5 for a correlation table. First, I examined the relationship between extraneous processing and performance on the five working memory tasks. None of the working memory task scores were significantly correlated with subjective extraneous processing when learning from a moderate distraction lesson. However, there was marginal significance between Corsi block task score and extraneous processing, $r = .23$, $p = .051$, such that an increase in working memory capacity is related to an increase in extraneous processing, which is not in the predicted direction. This also differed from the results found in Study 1a.

Table 5

Correlations between Cognitive Load and Working Memory Tasks in Study 1b

	Extraneous Processing	Essential Processing	Generative Processing
Stroop	-.03	.20	.03
Go/No-Go	.03	.26*	.12
N-back	-.12	.00	.21
Digit Span	-.07	.03	.03
Corsi	.23	.04	.03

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

Next, I examined the relationship between essential processing and performance on the five working memory tasks. There was a significant relationship between go/no-go task scores and essential processing, such that more errors on the task was related to more essential processing. This relationship demonstrated that those with weaker executive function reported more essential processing. This was once again a different finding than what was found in Study 1a.

Finally, I examined the relationship between generative processing and performance on the five working memory tasks. None of the tasks were significantly correlated with subjective generative processing. This was different from the findings in Study 1a.

Study 1c (Low Distraction Lesson)

The last study focused on understanding how learners' ability to manage and hold incoming information, measured specifically by executive function and working memory capacity tasks, related to learning from a lesson with a low level of distractions. Specifically, I examined the correlation between posttest scores and scores on executive function and working memory capacity tasks after learning from a lesson with few distracting features.

Method

Participants and Design. The study included 68 participants recruited from the Psychology Subject Pool at the University of California, Santa Barbara. The mean age of the participants was 19.09 years (*SD* = 1.58) and most of the participants were in their 1st or 2nd year at the university (*M* = 1.71 years, *SD* = .95). Additionally, the prior knowledge of the participants was somewhat low (*M* = 1.69, *SD* = .26) based on a 5-point Likert scale from "1 – Very Low" to "5 – Very High,". Of the sample, 40 participants identified as women and 28 participants identified as men. The predictor and outcome variables for this study were the same as Study 1a and Study 1b.

Materials. All of the materials were the same as Study 1a and Study 1b, except for the lesson.

***Lesson*.** The lesson participants watched contained the same topic material as the lesson presented in Study 1b, but without any distracting features introduced in the Study 1b video. The lesson was still an animation with a voice over, but only the visuals that contributed to the goal of the lesson were included in this version. The script was also modified to remove any of the seductive details, so participants only heard facts relevant to the lesson. This lesson was determined to be the least distracting of all the lessons because the seductive details in both the visual and verbal portion of the lesson were removedThe lesson lasted a total of three minutes and 45 seconds. Images from the lesson can be seen in Figure 1.

Postquestionnaire. The postquestionnaire was the same as Study 1a and Study 1b. The McDonald's Omega was .73 for extraneous processing questions, .64 for essential processing questions, and .49 for generative processing questions.

Apparatus. The apparatus used for this study were the same as in Study 1a and 1b.

Procedure. The procedure was the study as in Study 1a and 1b.

Results

Posttest Scoring. The posttest was scored in the same way as in Study 1a and Study 1b were scored. The correlation between the two scorers was strong, $r = .90$, $p < .001$. McDonald's Omega for the posttest was .65.

Outliers. Outliers were dealt with in the same way as they were in Study 1a and 1b. There were no outliers identified in this data.

Hypothesis 2: Do Posttest Scores Not Correlate with Scores on Executive Function and Working Memory Capacity Measures when Watching a Low Distracting Lesson? Hypothesis 2 predicted that posttest scores should not correlate with executive function and working memory capacity task scores as participants had watched a lesson that had a low number of distractions. The main analysis to test this hypothesis was that posttest scores would not correlate with scores on the executive function and working memory capacity tasks. As this lesson did not include many distractions, Hypothesis 2 was appropriate to test here. A correlation matrix was constructed to determine the relationships between posttest scores and each of the five working memory tasks, which is displayed in Table 6. None of the working memory tasks significantly correlated with posttest score,

Table 6

Correlation Matrix for Study 1c

	Posttest	Stroop	Go/No-Go	N-back	Digit Span	Corsi
Posttest	1.00	.16	-.19	.11	-.10	.13
Stroop		1.00	.18	.11	-.19	.05
Go/No-Go			1.00	.02	-.15	-.09
N-back				1.00	-.16	-.12
Digit Span					1.00	.15
Corsi						1.00

* $p < .05$, ** $p < .01$, *** $p < .001$
Lower scores on Stroop and go/no go tasks indicate better executive function skill

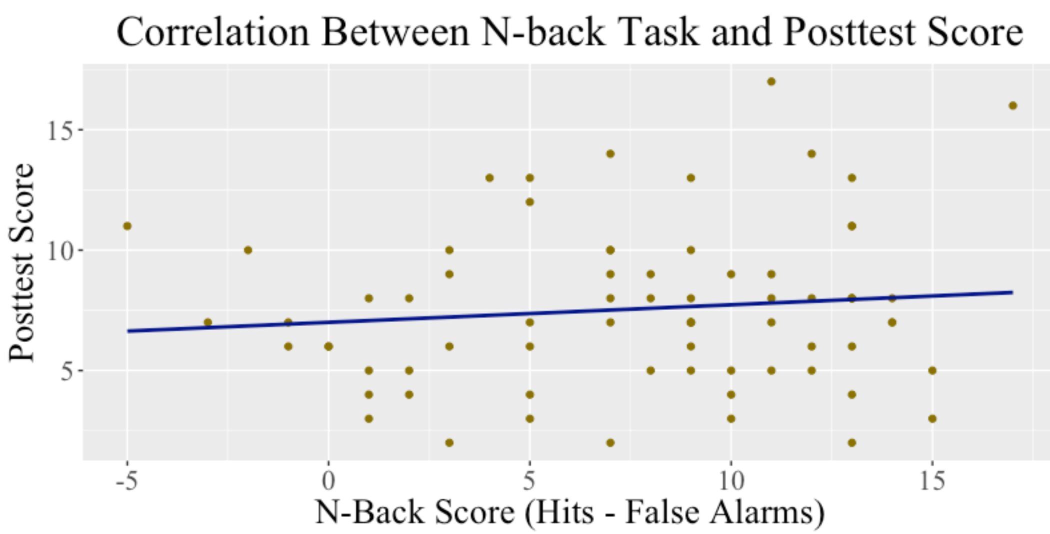
consistent with hypothesis 2. See Figure 4 for scatterplot of n-back scores and posttest performance, as a comparison to the two other studies in which n-back scores did correlate with posttest scores.

To supplement this analysis, a stepwise regression was used to determine if there was a predictive relationship between the working memory tasks and posttest performance for the low distraction lesson. All five predictors were included in the regression and the program selected only those that provided predictive power in the final output. The program did not select any of the variables to include in the stepwise regression, indicating that none of the predictors had predictive power over posttest score. This result is also consistent with hypothesis 2, suggesting that managing incoming information and holding information does not play a role in learning when participants do not encounter many distractions.

Exploratory Analyses. Several exploratory analyses were conducted to understand how executive function and working memory capacity interact with some of the

Figure 4.

Correlation Between N-back and Posttest Scores for Study 1c



postquestionnaire responses participants made. See Table 7 for a correlation table. First, I examined the relationship between extraneous processing and performance on the five working memory tasks. None of the working memory task scores was significantly correlated with subjective extraneous processing when learning from a low distraction lesson. Next, I examined the relationship between essential processing and performance on the five working memory tasks. Once again, none of the working memory score variables was significantly correlated with subjective essential processing when learning from a low distraction lesson. Finally, I examined the relationship between generative processing and performance on the five working memory tasks. Again, none of the working memory score variables was significantly correlated with subjective generative processing when learning

Table 7

Correlations between Cognitive Load and Working Memory Tasks in Study 1c

	Extraneous Processing	Essential Processing	Generative Processing
Digit Span	.05	.05	-.14
Stroop	.13	.00	-.08
Go/No-Go	.11	.02	-.13
Corsi	-.09	.02	.03
N-back	-.11	-.11	.00

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

from a low distraction lesson. These findings were in contrast to most of the corresponding findings in Study 1a and 1b.

Comparing the Three Studies

Are the Groups Equivalent on Key Characteristics?

Before proceeding with further analyses, I checked that random assignment was successful by comparing two conditions groups on basic characteristics. Means and standard deviations on basic characteristics for the two groups are presented in Table 8. There was not a significant difference between groups based on age, $F(2, 284) = .11$, $p = .900$; prior knowledge, $F(2, 297) = .05$, $p = .955$, or gender, $\chi^2(4, N = 299) = 8.08$, $p = .089$. However, there was a significant difference in class level, $F(2, 294) = 3.64$, $p = .028$, so it was included as a covariate in the mean comparisons.

Table 8.

Means and Standard Deviations for Comparisons Across Studies in Experiment 1

	High Distraction		Moderate Distraction		Low Distraction	
	M	*SD*	*M*	*SD*	*M*	*SD*
Age	18.88	3.24	18.89	1.15	19.04	1.55
Class Level	1.99	1.00	1.72	.89	1.67	.93
Prior Knowledge	1.70	.99	1.74	.85	1.69	.82
Extraneous Processing	2.95	1.01	3.12	1.03	3.30	.77
Posttest Performance	9.18	4.18	8.55	4.13	7.56	3.30

Hypothesis 3: Are the Correlations for Executive Function and Working Memory Capacity Tasks with Posttest Performance Different Across Lessons with Various Levels of Distraction?

Although this data was collected at different times, comparisons across the studies was of interest. One area of interest was comparing the correlations between n-back score and posttest performance. To compare the correlations, Fisher-z transformations were conducted. The *z'* scores for the correlation between n-back task score and posttest performance were .32 for the high distraction lesson, .27 for the moderate distraction lesson, and .11 for the low distraction lesson. The correlation in the high distraction lesson was not significantly different from the correlation in the moderate distraction lesson ($p = .711$) or from the correlation in the low distraction lesson ($p = .162$). Additionally, the correlation in the moderate distraction lesson was not significantly different from the correlation in the low distraction lesson ($p = .368$). These results are inconsistent with hypothesis 3. However, these comparisons are underpowered, which should be considered in interpreting the results presented here.

Hypothesis 4: Do Learners in Lessons with More Distraction Report Feeling More Distracted than Those in Lessons with Less Distractions?

Another comparison of interest was how levels of self-reported extraneous processing would vary across the different studies. Hypothesis 4 predicted that those who learn from lessons that have more distractions (i.e., the high or moderate distraction conditions) should report experiencing more extraneous processing than those who learn from lessons that have fewer distractions (i.e., the low distraction condition). This is because in a high or moderate distracting lesson, there are more distractions to pull attention away from information related to the objective of the lesson than the one without distractions. Mean and standard deviation for extraneous processing score is reported in Table 8. With regard to self-report extraneous processing, a one-way between-subjects ANCOVA, with class level as a covariate, was conducted and found that there was not a significant main effect of amount of distractions on learning, $F(2, 293) = 2.71$, $p = .068$. This was inconsistent with hypothesis 4.

Hypothesis 5: Do Learners in Lessons with More Distraction Learn Less than Those in Lessons with Less Distractions?

Hypothesis 5 predicted that those who learn from lessons that have more distractions should have worse learning, due to distractions pulling attention away from the relevant material, than those who learn from lessons that have fewer distractions. Means and standard deviations for posttest scores in the three conditions are reported in Table 8. Another one-way between-subjects ANCOVA, with class level as a covariate, was conducted for the posttest performance, and there was a significant main effect of amount of distractions on learning, $F(2, 293) = 3.45$, $p = .033$. Post-hoc Tukey test compared each group mean and found that those in the low distraction condition had significantly worse posttest performance than those in the high distraction condition ($p = .013$). None of the other comparisons were significant. This was inconsistent with hypothesis 5.

Discussion

Across all of these studies, we can see an important emerging pattern; for lessons that present distractions to the participants, i.e., both the high and moderate level of distraction lessons, learners' ability to manage incoming information was related to learning. Specifically, performance on the n-back task was significantly related to posttest performance for the lessons that presented distractions to learners (Study 1a and Study 1b). This suggests that as someone is better able to manage new information from a lesson, the better these learners are able to absorb the material. This is similar to what other research has found about the role that executive function plays in learning (i.e., Begolli et al., 2018; Meltzer & Krishnan, 2007; Rhodes et al., 2014; Zelazo et al., 2016). If an individual is better able to ignore information that is irrelevant and focus only on the relevant information, the more likely they will be able to learn the material presented in a lesson.

What is interesting about this research is that this pattern of results does not persist in all situations. The relationship between managing incoming information and posttest performance only persisted when the lesson had distracting features (Study 1a and 1b); the pattern of results went away when a lesson had low levels of distraction (Study 1c). Although these findings are highly intuitive, they have not been established previously in research. As such, these findings set up an important pattern for investigating distraction in other learning environments. These results also demonstrate that there are learning environments that can be beneficial for all learners, regardless of their executive function; in environments that do not include distracting material, it seems as though all learners are able to engage with the material, without issues arising from their ability to manage incoming information.

This research also suggests that how much information someone can hold at one time does not seem to relate to learning, or at least not in the same way that managing incoming information does. Across all three lessons, working memory capacity did not correlate with posttest performance. As such, it seems as though there is an important distinction between the roles that managing information and holding information play in learning; these two different components of working memory do not have the same impact on learning. As such, it may be most important to focus on how well learners can manage incoming material through effective cognitive processing.

Additionally, the exploratory findings regarding the relationship between working memory tasks and different types of cognitive load were conflicting throughout the studies. These conflicting findings make it difficult to draw any conclusions about the role that managing information and holding information may play in different types of cognitive load. A potential reason for the variability in the relationships between these characteristics may point to the fact that these measures of cognitive load are unreliable. However, as these were exploratory findings and not the main aim of this research, this will not be discussed further.

When directly comparing the correlations across studies, there were no differences. This is not what is expected, but one issue with this research is that each study was conducted at different times and some of the studies may not have enough participants to provide sufficient statistical power. As such, more research needs to be conducted, preferably in a study that directly compares lessons with various amounts of distraction, to understand whether the correlation between managing incoming information and posttest performance is significantly different at various levels of distraction.

Lastly, the mean comparisons across the three different studies presented interesting, but conflicting results. In comparing the three studies, for both self-report extraneous processing and learning, there were significant differences with the low distracting study demonstrating higher extraneous processing and worse learning. These results are hard to explain, but one potential suggestion may be that having irrelevant, but interesting facts in a lesson may be helpful for learners to continue to pay attention during the lesson as opposed to having their mind wander. Further research is needed to understand why these results occurred.

Chapter III: Experiment 2

Rationale and Hypotheses

The goal of Experiment 2 was to extend the results of Experiment 1 and understand how individual differences in how well learners manage incoming information and how much information learners can hold interacts with learning from learning environments that have potential for distractions. The learning environment that was of particular interest in this research was immersive virtual reality (IVR). IVR is a technology that can introduce learners to various distractions due to the fact that information is presented 360-degrees around a learner at all times in the device, which is much more information than would be presented in a traditional learning environment, like a video lesson on a desktop computer screen. Learners may be overwhelmed by the perceptual richness of the environment (Mayer et al., 2022). Thus, IVR has potentially distracting features that a learner would need to deal with to learn.

The main question being addressed with this research was: does ability to manage incoming information and capacity for new information relate to learning from different learning media? More specifically, this researched aimed to understand whether executive function and working memory capacity correlate with learning from both a potential distracting learning technology, IVR, and from a more traditional and less distracting learning technology, a slideshow. A secondary question for this research was how well individuals learn in an IVR lesson compared to a slideshow lesson. This second question was interested in testing the coherence principle as learning in a slideshow lesson would follow this principle more than an IVR lesson. I investigated this question by designing a one-way between-subjects design in which participants received the same material either in an IVR lesson or in a slideshow lesson, and then completed a posttest and tasks to measure working memory factors.

As with Experiment 1, the first hypothesis predicted that learners who are better at managing incoming information should learn more information from a lesson with a high level of distraction (i.e., IVR) but not from a lesson with a low level of distraction (i.e., slideshow). More specifically, it was predicted that posttest scores would correlate with scores on executive function tasks when a lesson was presented in IVR (hypothesis 1a), but there should not be a correlation when a lesson was presented in a slideshow (hypothesis 1b). The second hypothesis predicted that the amount of information a learner can hold while learning should impact learning when distractions are present. More specifically, it is predicted that posttest scores would correlate with scores on working memory capacity tasks in learning from an IVR lesson (hypothesis 2a), but not in learning from a slideshow lesson (hypothesis 2b). Lastly, if IVR does present more distractions to learners than a slideshow lesson, it was predicted that learners in the IVR lesson would report higher subjective experiences of distraction (hypothesis 3a) and perform worse on a posttest (hypothesis 3b) than learners watching a slideshow lesson.

Method

Participants and Design

The participants were 161 participants recruited from the Psychology Subject Pool at the University of California, Santa Barbara. However, 11 participants did not complete the study, so their data were not included in any analyses, which left 149 participants with complete data. The mean age of the remaining participants was 19.76 years (*SD* = 3.07) and most of the participants were in their 1st or 2nd year at the university (*M* = 2.00 years, *SD* = 1.17). Additionally, the prior knowledge of the participants was somewhat low (*M* = 2.13, *SD* = .94) on a 5-point Likert scale from “1 – Very Low” to “5 – Very High.” Of the sample, 109 participants identified as women, 36 participants identified as men, and three participants identified as non-binary. In a between-subjects design, there were 74 participants randomly assigned to the IVR condition and 75 participants randomly assigned to the slideshow condition. For the first set of analyses, the predictor variables were executive function task scores, including scores on the Stroop task, n-back task, and flanker task, and working memory capacity task scores, including scores on the digit span task, Corsi block task, and operation span task. The outcome variable was posttest score. For the second set of analyses, the independent variables were the condition participants were in, self-reported distraction rating, and subjective extraneous processing rating. The dependent variable was posttest performance.

Materials

All the materials were computer-based or presented via a head-mounted display and were presented in a controlled laboratory setting. They included a prequestionnaire, lesson, posttest, postquestionnaire, and working memory tasks.

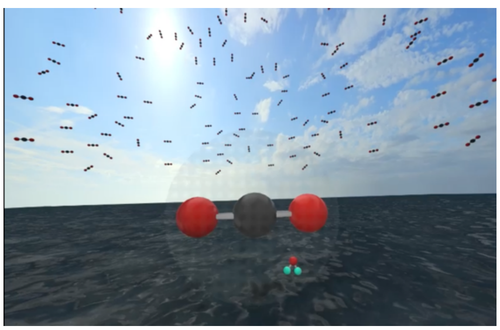
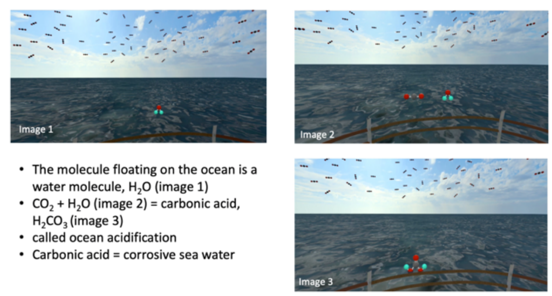
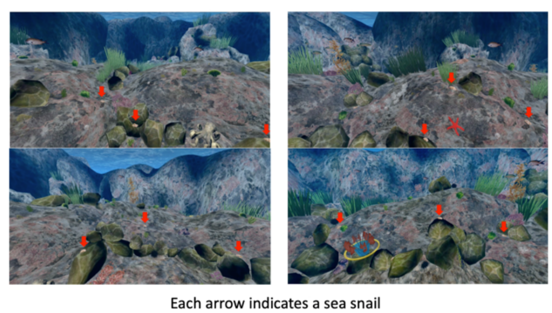
Prequestionnaire. The prequestionnaire solicited information about the basic characteristics of the participants, including their age, gender, class level, and prior knowledge. Prior knowledge on ocean acidification was assessed on a 5-point Likert scale from “1 – Very Low” to “5 – Very High.” As with Experiment 1, this was used to gather information about each participant's subjective prior knowledge of the topic instead of using a pretest in order to not introduce testing effects into the study. Additionally, the prequestionnaire assessed for participants' proneness to motion sickness. Only one participant reported being extremely prone to motion sickness, but that participant had been randomly assigned to the slideshow condition, so no participants with full sets of data had to be excluded from the study. Appendix A shows the prequestionnaire for this study.

Lesson. The lesson, in both IVR and slideshow, displayed information about what ocean acidification is, how it occurs, and how this process can harm sea life. In a series of six scenes, participants learned information about a main source of carbon dioxide (i.e., cars), how carbon dioxide and water interact to create carbonic acid, what a healthy reef looks like, what an acidified reef looks like, and how to prevent ocean acidification in the future. This lesson was either presented in immersive virtual reality or with a slideshow lesson. Images from these lessons are displayed in Figure 5.

Immersive Virtual Reality Lesson. The lesson used in IVR came from an openly accessible lesson called "The Stanford Ocean Acidification Experience" and was created by the Virtual Human Interaction Lab at Stanford University (Fauville et al., 2021). In each scene of the lesson, participants were fully immersed in the environment and were able to interact with the lesson, like 'feeling' carbon dioxide bubbles coming up from the sea floor or placing marker flags next to healthy sea snails. The information was narrated to the students to guide them through the lesson.

Figure 5.

Images from the IVR and Slideshow Lessons in Experiment 2

IVR Lesson	Slideshow Lesson
	
	
	

The first scene placed participants in a dark room facing a globe as the narration discussed how the world is changing due to human impacts. Then, participants were moved to the middle of a busy road with cars surrounding the participant to learn about how carbon dioxide comes from car exhaust and goes into the air. Then, participants were transported to a boat floating in the ocean. Here they could see the carbon dioxide molecules floating in the air and saw how these molecules interact with water molecules to create carbonic acid.

Next, participants were taken under the ocean and first shown a real image of a reef off the coast of Italy that is in danger of dying due to the effects of ocean acidification. Then, the participants moved into a computer-generated reef that was healthy and the narration discussed how marine scientists can identify if a reef is healthy by making different observations about the reef, like a species count. While in this environment, the participant was asked to place flags next to sea snails to simulate a species count. To contrast this environment, the participant was then moved into a part of the reef that had carbon dioxide vents coming from the sea floor. In this environment, participants were asked to once again do a species count of the sea snails in order to notice that there are no sea snails in this environment. Finally, the participant was transported back to the first scene where they saw the globe while the narration discusses what actions they can take to reduce the impacts of ocean acidification. The lesson lasted a total of seven minutes. The script for the lesson is included in Appendix D.

This IVR lesson could be considered more distracting than a lesson presented in a more conventional way for several reasons. First, the IVR lesson presents much more information to look at, much of which is not specifically relevant for the participant to look at to understand the material of the lesson. For instance, when watching how H_2O and CO_2 combine, in the IVR lesson, participants can turn around and see CO_2 molecules hanging around in the air or look at the boat they are standing on. These features will not help them learn the material but are present in the lesson to make the learner feel immersed in an entirely new environment.

Slideshow Lesson. The slideshow lesson was created using images and the script from the IVR lesson. Screenshots were taken from the IVR lesson and put in a slideshow with text highlighting the important information underneath the images. The same narration used in the IVR lesson was used in this lesson, except the parts that encouraged participants to engage with the environment were modified to fit the slideshow lesson. The participants were shown the same scenes as the IVR lesson, but the scenes were translated into still images and bullet points. There were 14 slides that translated the same information from the IVR lesson in this slideshow lesson. This lesson lasted just under five minutes. The script for this lesson is provided in Appendix D and slides are included in Appendix E.

The slideshow was considered not as distracting as the IVR lesson because it only presented the material that was relevant to the lesson. Thus, to extend the example about when H_2O and CO_2 combined, in the slideshow lesson, only the images of this interaction were shown; participants were not shown other images in the environment.

Posttest. The posttest consisted of eight open-ended questions to assess participants understanding of the material presented in the lesson. All of the questions aimed to assess transfer of knowledge from the lesson at some level but varied in how much transfer participants needed to engage in to successfully answer the question. The questions were: (1) "Explain how ocean acidification occurs." (2) "If car's exhaust released a different greenhouse gas, like methane (CH_4), could ocean acidification still occur due to car fumes?" (3) "What steps would need to be taken to make the ocean less acidic?" (4) "What molecules combine to increase ocean acidification? What molecule do they create?" (5) "How does an abundance of CO_2 in the atmosphere affect sea life?" (6) "Why were there snails in the first area of the reef but not in the second area?" (7) "How might ocean acidification affect humans?" (8) "How can marine scientists figure out if an area of a reef is healthy or not?" The posttest questions were presented to the participants in the same way that they were in Experiment 1.

Postquestionnaire. The postquestionnaire included 33 items about the participants' experience with the lesson. The first five items, which were the same as those in Experiment 1, assessed participants' experience with the lesson. The next nine items assessed the participants' subjective cognitive load experience and were the same as those from Experiment 1 (Parong & Mayer, 2018, 2021a, 2021b). The McDonald's Omega was .76 for extraneous processing questions, .55 for essential processing questions, and .62 for generative processing questions.

The next seven questions assessed the level of presence that participants felt while learning in the study, including items: "I felt that I was able to control events in the environment." and "I felt like I was really in the environment." on a 5-point Likert scale from "1 – Strongly Disagree" to "5 – Strongly Agree" (Parong & Mayer, 2018, 2021a, 2021b). McDonald's Omega for this set of questions was .91. For those only in the IVR condition, the next nine items assessed for cyber sickness brought on by the IVR experience (Kim et al., 2018). These items included two subscales to ask how much participants experienced oculomotor issues, such as fatigue and eye strain, and disorientation issues, such as headache and vertigo. Participants responded on a 4-point Likert scale from "1 – Not

At All" to "4 – Very." McDonald's Omega for the oculomotor questions was .66 and for the disorientation questions was .63. The final three items assessed how much experience participants had with VR prior to participating in the study, asked for participants to report any additional comments about the experience, and gathered information about if anything went wrong during the study. Postquestionnaire items are included in Appendix B.

Working Memory Tasks. A similar battery of working memory tasks from Experiment 1 was used to assess participants ability to manage incoming information (executive function) and ability to hold new information (working memory capacity). For this experiment, there were several changes to the tasks participants engaged in; the go/no-go task was removed and the flanker task and the operation span task were added.

In Experiment 1, it was concluded that participants were too good at completing the go/no-go task and rarely made errors. This made the distribution of the scores on this task very limited and not a good measure of individual variability. For this reason, the go/no-go task was not used in Experiment 2. The flanker task was added as a replacement measure of executive function, as it has been used in previous research of measure individual differences in executive function (Diamond, 2012; Eriksen & Eriksen, 1974).

Additionally, the operation span task was added because some literature suggests that simple span tasks, such as the digit span and Corsi block task, are not assessing working memory capacity as they do not predict performance on tasks that would relate to working memory capacity (Redick et al., 2012). The reason that the operation span task may be a better measure of working memory capacity is because it measures how much participants are able to hold in working memory *while* processing other types of information (Daneman

& Carpenter, 1980; Ilkowska & Engle, 2010), which is much more similar to how working memory capacity is used in learning situations (Engle et al., 1999). As such, this task was added to determine if a task that assess capacity while working with other information plays a role in learning. Thus, participants completed six classic working memory tasks to help assessed individual differences in executive function and working memory capacity.

The digit span task was presented on Qualtrics and each sequence length was completed twice. Stroop, Corsi block, n-back, and flanker tasks were presented on PsyToolKit (Stoet, 2010, 2017) and each task was completed twice. The first time the participants completed the task served as a training block then the second time they completed the task served as the real block. Only the real block responses counted for the participants' scores on each task. Lastly, participants completed the operation span task using Inquisit. Inquisit is a platform run by Millisecond that administers cognitive tests and assessments.

Stroop Task. The Stroop task was the same as the one in Experiment 1.

N-back Task. The n-back task was the same as the one in Experiment 1.

Flanker Task. On the screen, participants were told they would see a target in the middle of several other images. The target would either point to the left ("<") or to the right (">") and participants had to respond with "A" if the point was to the left or "L" if the point was to the right. In trials that were congruent, the surrounding arrows would point in the same direction (e.g., "<<<<<"). In trials that were incongruent, the surrounding arrows would point in the opposite direction (e.g., "<<><<"). Before the trial began, participants would first see a fixation cross. About 1/3 of the trials were congruent and the other 2/3 of the trials were incongruent. Participants completed 50 trials within each the practice and real blocks. For this set of studies, the score for the flanker task was determined by subtracting the averaged reaction time of the congruent trials from the averaged reaction time of the incongruent trials. Lower scores indicated better executive function.

Digit Span Task. The digit span task was the same as the one in Experiment 1.

Corsi Block Task. The Corsi block task was the same as the one in Experiment 1.

Operation Span Task. On the screen, participants were given two practice tasks to practice each component of the operation span task before completing the real task. The first part of the practice was dedicated to practicing the letter sequence portion of the task. One at a time, participants were shown letters in a sequence of three to seven letters. Then, once the sequence was complete, participants would see a grid of letters that they would have to click on to report the sequence they just saw. In the second part of the practice, participants practiced the math portion of the task. First, they were shown an equation, such as (2*1) + 3 = 7. Participants were told to click the mouse to move forward. On the next screen, participants were asked to report whether the equation was true or false. Participants were told to keep their accuracy rate for the math problems above 80% as a way to ensure they were putting effort into this task. Once both practice portions were complete, participants were taken to the real task in which they were asked to report whether or not the displayed equation was true or false then shown a letter. This sequence repeated until they were shown three to seven letters after completing math equations. Once the sequence was complete, participants were shown a grid of letters and told to report the sequence they saw by clicking on the letters in the correct sequence. Participants completed 15 sequences in the real trial after they did three practice sequences. For this set of studies, the score for the operation span task was the number of letters they reported correctly in fully correct reported sequences. Higher scores indicated higher working memory capacity.

Apparatus

For the IVR lesson, participants used the HTC Vive head-mounted display virtual reality system connected to a Dell Alienware computer. The lesson was operated through Steam software on the Dell computer. In addition to the head-mounted display, participants had wireless controllers that they used to interact with the environment and receive haptic feedback from the virtual environment. The slideshow lesson and all the non-lesson material were presented on a Dell computer with a 20-inch color monitor.

Procedure

Participants started the study by signing the consent form. Once they agreed to participate, they completed the prequestionnaire, the digit span task, and the operation span task. Then, they were randomly assigned to either the IVR lesson or the slideshow lesson and watched the appropriate lesson. For those in the IVR lesson, they were instructed on how to put on the head-mounted displayed and monitored while wearing the headset. For those in the slideshow condition, they were instructed to use headphones to watch the lesson and were monitored while watching the lesson. When they had completed the lesson, participants were thanked for their participation and told to return for the second part of the study a week later.

In the second session, participants started by completing the posttest. Once done with the posttest, the survey moved directly to the postquestionnaire questions. Then, participants were instructed to complete the Stroop task, Corsi block task, n-back task, and flanker task, each of which they completed twice in a row. Once done with that, participants were thanked again for their participation and given credit for their participation. Each session of the experiment took no longer than an hour to complete. The study was approved by the university's Institutional Review Board (IRB).

Results

Posttest Scoring

The posttest was scored with a rubric that assessed how many main points were present in each of the participants' responses to the posttest questions, similar to how the posttest in Experiment 1 was scored. For example, a question like "How can marine scientists figure out if an area of a reef is healthy?" had three main points associated with it and that could earn the participants points. Specifically, these main points were "conduct a species count," "if there is more sea life, the reef is healthier," and "if there is no or little sea life, the reef is unhealthy." As such, if participants did not earn all the points, this did not necessarily indicate that their answer was incorrect, but instead indicated how many of the relevant and correct details were included in their response to the question. Participants could respond to a question correctly (e.g., in the example question, a correct answer would say something like "conduct a species count.") but not include all the main points in the rubric for that item. This rubric style was used in order to determine who demonstrated more knowledge about the material by including more specific facts to support their response rather than a binary right/wrong grading scheme. The main points associated with each question are provided in Appendix B.

Each question had two to seven main points associated with it. When a participant's response included the main idea from the rubric in their answer, the response was given a "1" for that specific rubric item, and if not, it was given a "0." Then a total score for each question was created by adding the 1s and 0s across the rubric. Then, a total score for the posttest was created by adding the total scores for each question, to create a number out of 32. The rubric is provided in Appendix B.

Two independent scorers read each response and gave the participants scores. The correlation between the two scorers was strong, $r = .86$, $p < .001$. All differences in scoring between the scorers was discussed until 100% agreement. The reliability of the posttest was determined using McDonald's Omega and was .78.

Outliers

In this data, outliers were determined to be participant scores on the measures of working memory in which the z-score exceeded positive or negative three. In this data set, four participants were removed from the data because of their outlier scores, including one participant due to their high Stroop task score, one participant due to their low Corsi block task score, one participant due to their high flanker task score, and one participants due to their low n-back score. Removing these outliers did not change the pattern of results.

Are the Groups Equivalent on Key Characteristics?

Before proceeding with further analyses, I checked that random assignment was successful by comparing two conditions groups on basic characteristics. Means and standard deviations on basic characteristics for the two groups are presented in Table 9. There was not a significant difference between groups based on age, $t(140) = -.58$, $p = .563$; class level,

Table 9

Means and Standard Deviations for Experiment 2

	IVR Condition		Slideshow Condition	
	M	*SD*	*M*	*SD*
Age	19.66	2.47	19.97	3.69
Class Level	2.06	1.24	2.04	1.12
Prior Knowledge	2.13	.96	2.16	.94
Distraction from Lesson	1.79	.98	1.91	.83
Extraneous Processing	2.21	.82	2.50	.90
Posttest Score	11.88	4.71	11.79	5.23
Presence	4.17	.60	2.87	.70
Enjoy	4.53	.53	3.86	.73
Interesting	4.44	.60	4.06	.83
More Lessons	4.33	.63	3.73	1.01
Effective	4.28	.81	3.40	.97
Motivated	4.51	.67	3.70	.97

$t(140) = .06$, $p = .949$; subjective prior knowledge, $t(140) = -.20$, $p = .841$; or gender ratios, χ^2 (2, $N = 144$) = .46, $p = .794$. As such, it was determined that random assignment was successful in creating equivalent groups for these key characteristics.

Hypothesis 1: Do Posttest Scores Correlate with Scores on Executive Function Tasks when Learning via IVR or Slideshow?

Hypothesis 1 predicted that posttest scores should correlate with executive function task scores for participants who watched a lesson in IVR, as IVR lessons introduce distraction into the learning environment. Hypothesis 1 also predicted that posttest scores should not correlate with executive function task scores for participants who watched a slideshow lesson, as this lesson should not introduce many distractions into the learning environment. The main analysis to answer this hypothesis was to correlate posttest scores with scores on executive function tasks for each condition. Thus, the data first was split to separate the two conditions from one another. For each condition, a correlation matrix was constructed to determine the relationships between posttest scores and the three executive function tasks, which is presented in Table 10. For the IVR lesson, the flanker task scores were significantly correlated with posttest scores, $r = -.23$, $p = .049$, partially in line with hypothesis 1a. See Figure 6a for scatterplot of flanker task and posttest scores. None of the other executive function tasks was significantly correlated with posttest scores for the IVR lesson. For the slideshow lesson, none of the executive function tasks scores was significantly correlated with posttest score, in line with hypothesis 1b.

To supplement these analyses, a stepwise regression was used for each condition to determine if there was a predictive relationship between the executive function tasks and posttest performance. The three predictors were included in the regression and the program selected only those that provided predictive power in the final output. For the IVR lesson, only the flanker task had predictive power for the posttest scores, so it was the only variable of the three that was included in the stepwise regression, $R = .23$, $R^2 = .05$, $F(1, 72) = 4.00$, $p = .049$. The flanker task was related to posttest performance in the predicted direction, such that as flanker score decreased–indicating an increase in executive function–posttest scores increased, $B = -.01$, $\beta = -.23$, $t = -2.00$, $p = .049$. For the slideshow lesson, none of the variables was selected by the program, indicating that none of these variables had predictive power over posttest scores. Overall, there is some evidence that executive function is related

Table 10

Correlation Matrix for IVR and Slideshow Lessons in Experiment 2

	Posttest	Stroop	N-back	Flanker	Digit Span	Operation Span	Corsi
IVR Lesson							
Posttest	1.00	-.12	.19	-.23*	.05	.14	.02
Stroop		1.00	-.09	.18	-.33**	-.31**	.04
N-back			1.00	-.41***	.30**	.37**	.27*
Flanker				1.00	-.34**	-.27*	-.06
Digit Span					1.00	.59***	.10
Operation Span						1.00	.02
Corsi							1.00
Slideshow Lesson							
Posttest	1.00	-.04	.03	.07	.16	.19	-.08
Stroop		1.00	.19	.21	-.05	-.05	.05
N-back			1.00	-.13	.23*	.10	.05
Flanker				1.00	-.21	-.20	-.08
Digit Span					1.00	.50***	.19
Operation Span						1.00	.26*
Corsi							1.00

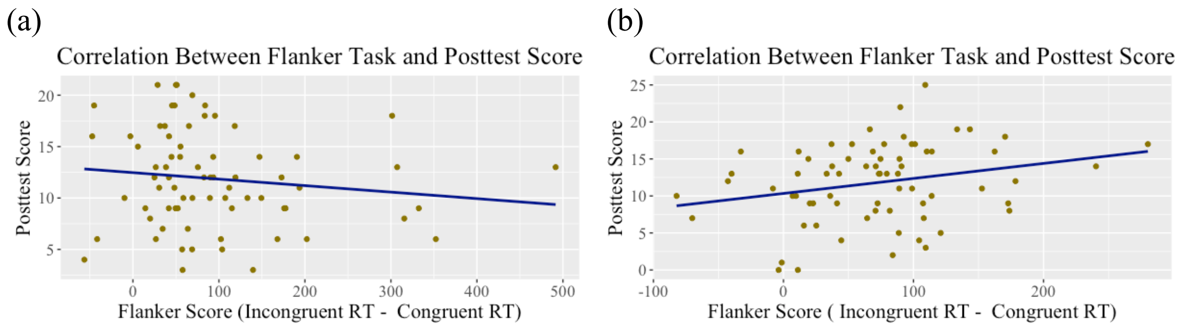
* $p < .05$, ** $p < .01$, *** $p < .001$

to learning outcomes when students learn with a distracting medium (i.e., IVR) but not when they learn with a less distracting medium (i.e., slideshow).

As a secondary analysis to compare the relationship between flanker task scores and posttest performance across the two conditions, a moderation analysis was conducted. In this analysis, flanker score was the predictor variable, posttest performance was the outcome variable, and lesson type was the moderator. See Figure 7 for a graph of the moderation. The moderation analysis demonstrated that there was not a significant interaction between flanker score and lesson type of learning, $p = .169$. This suggests that, although the correlation between flanker task score and posttest performance for the IVR lesson was

Figure 6.

Correlation Between Flanker Task and Posttest Scores for Experiment 2



(a) Correlation between flanker task and posttest performance for the IVR condition (b) correlation between flanker task and posttest performance for slideshow condition

significant and it was not for the slideshow lesson, this difference was not large enough to be statistically significant.

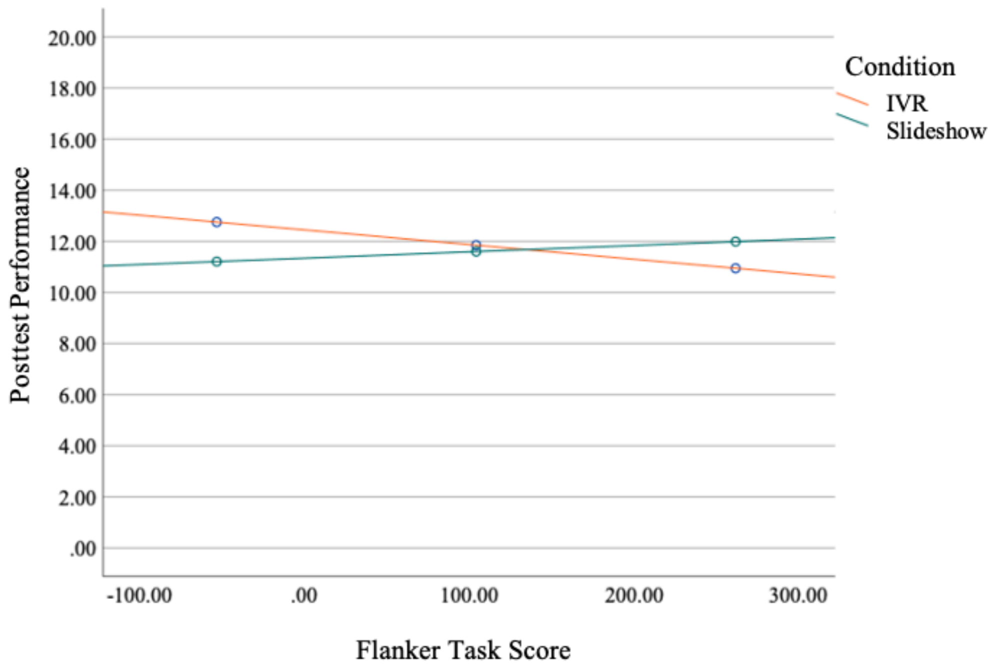
Hypothesis 2: Do Posttest Scores Not Correlate with Scores on Working Memory Capacity Tasks When Using IVR or Slideshow Lessons?

Hypothesis 2 predicted that posttest scores should correlate with working memory capacity task scores. The main analysis to answer this hypothesis was to correlate posttest scores with scores on working memory capacity tasks. A correlation matrix was constructed to determine the relationships between posttest scores and the three working memory capacity tasks, which is presented in Table 10. For both the IVR lesson and slideshow lesson, none of the working memory tasks scores was significantly correlated with posttest score, which is inconsistent with hypothesis 2a but consistent with hypothesis 2b. See Figure 6b for the scatterplot of flanker task and posttest scores.

To supplement these analyses, a stepwise regression was used for each condition to determine if there is a predictive relationship between the working memory capacity tasks

Figure 7.

Moderation Analysis for Flanker Score Task on Posttest Performance By Condition



and posttest performance. The three predictors were included in the regression and the program selected only those that provided predictive power in the final output. For both the IVR lesson and the slideshow lesson, none of the variables was selected by the program, indicating that none of these variables had predictive power over posttest scores. Overall, hypothesis 2a was not supported, but hypothesis 2b was supported.

Hypothesis 3: Do Learners in an IVR Lesson Experience More Distractions and Worse Learning than in a Slideshow Lesson?

Hypothesis 3 predicted that those who learn a lesson in IVR should experience more distractions in a lesson and thus should report more distractions and extraneous processing than those who learn in a slideshow lesson. Hypothesis 3 also predicted that those who learn a lesson in IVR should have worse learning compared to those who learn from a slideshow lesson. To test this hypothesis, posttest score, responses to the question about how distracted participants felt during learning, and responses to the cognitive load questions focused on extraneous processing were analyzed using *t*-tests. Means and standard deviations of these variables from these two groups are reported in Table 9.

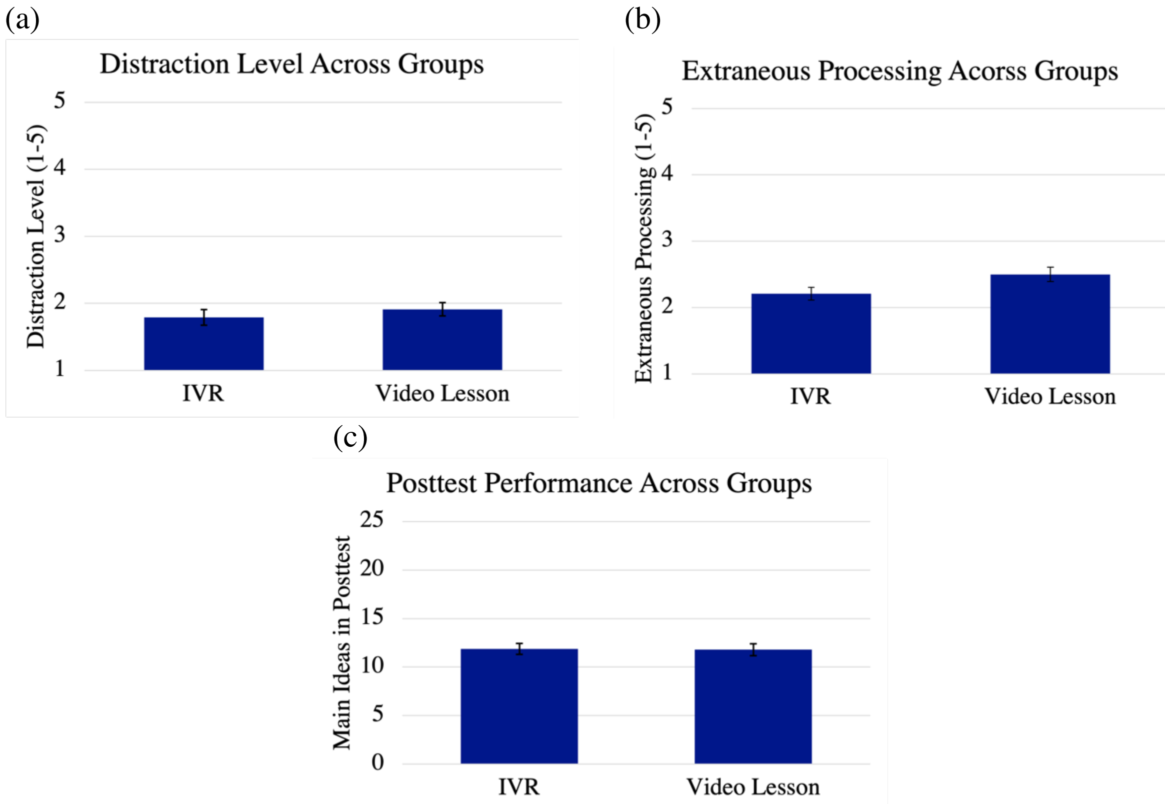
First, for how much participants felt distracted during the lesson, there was not a significant difference between those who learned in the IVR lesson and those who learned in the slideshow lesson, $t(140) = -.81$, $p = .422$. See Figure 8a for bar graph of distraction across groups. For the amount of extraneous processing participants reported feeling, there was not a significant difference, but the difference was trending toward significant, $t(140) = -1.96$, $p = .052$, with those who learned in the slideshow lesson reporting higher levels of extraneous processing than the IVR lesson. The lack of differences on the distraction and extraneous processing questions contradicted hypothesis 3a. See Figure 8b for bar graph of extraneous processing across groups. Next, for posttest performance, there was not a significant difference between those who learned in the IVR lesson and those who learned in the slideshow lesson, $t(140) = .11$, $p = .915$. This finding does not support hypothesis 3b. See Figure 8c for bar graph of posttest scores across groups.

Exploratory Analyses

Several other analyses were of interest to this research, but not specifically predicted in the hypotheses. Means and standard deviations on the ratings of the participants' perceptions of the lesson are provided in Table 9. First, sense of presence was compared

Figure 8.

Mean Differences Across Conditions in Distraction and Posttest for Experiment 2



(a) subjective experience of distraction across VR and slideshow condition, (b) subjective experience of extraneous processing across VR and slideshow condition, (c) posttest scores across VR and slideshow condition

between the IVR and the slideshow conditions, as the IVR condition should experience more presence during learning than the slideshow lesson. As expected, those in the IVR condition reported more presence while learning than those in the slideshow condition, $t(140) = 11.92, p < .001$.

Previous research has also demonstrated that students feel more positively about learning with IVR, compared to more traditional lessons. As such, ratings of enjoyment, interest, desire to learn similarly in the future, subjective effectiveness, and motivation were compared between the IVR and slideshow conditions. Participants in the IVR condition, compared to the slideshow condition, reported that they enjoyed the lesson more, $t(140) = 6.29$, $p < .001$, were more interested in the lesson, $t(140) = 3.18$, $p = .002$, desired more similar lessons in the future, $t(115) = 4.28$, $p < .001$, felt the lesson was more effective for their learning, $t(134) = 5.85$, $p <. 001$, and were more motivated to learn, $t(122) = 5.81$, $p < .001$. It appears that participants who learned with the IVR lesson had a better and more motivating learning experience compared to those who learned with the slideshow.

Comments on Lesson. Open-ended comments about the experience were also encouraged, and as such, participants' comments are briefly analyzed here. All comments were coded for different themes that were present in their responses. To begin, there were many more comments submitted when the participants learned with the IVR lesson (31 comments) compared to when participants learned with the slideshow lesson (16 comments). The nature of the comments was similar in some ways, but also different. For example, how positively participants felt about the lesson was a common theme across both conditions but was more often commented after the IVR condition (13 comments – 41.93%) compared to the slideshow condition (4 comments – 25.00%). For participants in the IVR condition, they made comments like "I liked it very much." "enjoyed the lesson, overall positive experience," and "Very fun. Would be interested in doing it again with my course material." For participants in the slideshow condition, they made comments like "Overall the lesson including videos were interesting," "I enjoyed it, but I did not notice much special about how it was taught." and "I enjoyed the lesson on ocean acidification."

Another theme that came up across both conditions was how engaging the lesson was. Once again, engagement was discussed much more in the IVR condition (11 comments – 35.48%) compared to the slideshow condition (2 comments – 12.50%). The comments for the IVR condition were overwhelmingly positive and included things like "I found it engaging.” “I liked the interaction of the lesson.” and “I think the idea of teaching lessons through VR is a really cool idea and I liked the interactive aspect of it.” The comments for the slideshow condition were more mixed, with a positive comment like “It was really interactive,” but a negative comment like “I think one of the biggest things that made the less engaging was the limited graphics.”

One important theme that came up across both conditions was the distraction participants felt while learning. Once again, there were more instances of comments about distraction in the IVR condition (12 comments – 38.71%) than in the slideshow condition (5 comments – 31.25%). However, more specifically, the types of distractions that participants reported were different across the two conditions. In the IVR condition, one type of distraction participants discussed was the visuals, such as “The visuals of the VR was very blurry, difficult to focus.” and “The VR does cause some eye strain because the screen is close to my eyes.” Another theme that came up for the IVR lesson was that there was a lack of text in the lesson to help participants follow along, like “one problem for me with learning with VR is that when I watch a video, I can always turn subtitle on, and I think I can learn better with text presented” and “I would love to see subtitles for the main points being made in the lesson. I feel like the voiceover is a little fast especially in the end of the lesson.”

An additional theme that came up for the IVR lesson was participants being distracted by being uncomfortable in some way, for example “Outside of being self-conscious that other people were watching me while I was doing the VR lesson, the experience itself was really cool.” and “It made me spin for a second, but then my eyes and body adjusted to the environment.” Lastly, a theme regarding the novelty of VR came up, such as “I’ve never used a VR headset before so I was excited to try it out with the lesson but I found myself a little distracted by the technology and was unable to focus entirely on the lesson at hand.” and “it was the first time I had ever done VR, so it was a little mesmerizing. I think the more I worked with VR lessons, the more effective they may become.”

The slideshow condition participants also made comments about distraction, but the themes that came up were slightly different. The main distraction theme came from being distracted by the previous task, such as “I was more focused on the math and letter memorization aspect so when it came to thhe video my mind was so exausted.” and “The pre-activity was extremely mentally draining, so I felt like I couldn’t focus on the ocean acidity material because I was so tired.” There was also one comment about the pace of the slideshow lesson being distracting, specifically “The lesson was relatively slow-paced, and this made it a bit harder for me to concentrate. But knowing that I may be tested on the material, I paid attention.”

Across these comments, it seems as though those in the IVR condition were more likely to report about their positive experience with the lesson, suggesting that IVR is seen as an intriguing learning tool for many students. That being said, IVR also seemed to cause more distraction for the participants than the slideshow condition, including from unfamiliarity with the technology itself. For the slideshow lesson, it seemed that it could cause distraction, but mainly from participants losing interest in the material. This suggests that there may be different sources of distraction across different media that should be assessed differently.

Discussion

The main finding of Experiment 2 focuses on the role that executive function plays when learning from an IVR lesson compared to learning from a slideshow lesson. Experiment 1 demonstrated that executive function was correlated with learning, but only when distractions were present in the lesson. This finding can be used to help understand learning in this IVR lesson. In this experiment, executive function played a role in learning from an IVR lesson, but not in learning from a slideshow lesson. As such, this suggests that there is something about learning with IVR that is distracting learners. However, the moderation analysis not showing an interaction between flanker task and type of lesson was unexpected, making the results a little harder to draw concrete conclusions from.

The comments from the lessons demonstrated that learning in IVR seemed to be more exciting and more motivating to students than the slideshow lesson. However, it also seemed as though the IVR lesson introduced more types of distraction that the learners had to contend with compared to the slideshow lesson. This result matches well with the correlational findings from the flanker task and posttest performance relationship. However, even though the reliance on executive function is suggesting that an IVR lesson introduces distraction into a lesson, it does not seem to impact group mean learning outcomes compared to a slideshow lesson.

Additionally, this study found some peculiar results, namely, the finding that there were no differences in how distracted participants found the lessons and how much extraneous processing each group engaged in. There was also no difference found in posttest scores between the two groups. These findings contradict the hypotheses, the coherence principle, and previous research on cognitive load induced by IVR lessons, and they also contradict the reliance on executive function that was present when learning with IVR but not with a slideshow lesson. This may suggest that the questions being asked are not focused enough on the issue or that the questions are not specifically designed in a way that makes learners think about all sources of distraction within an IVR lesson.

For example, one of the extraneous processing questions askes participants to agree or disagree if their mind was focused on something other than the lesson. Although this question seems straightforward enough, in an IVR lesson, it might not actually be assessing whether the participant was distracted or not. In an IVR lesson, a source of distraction can come from attention being pulled away from the important part of the lesson, but the participant can still focus their attention somewhere else within the lesson. As such, these participants would likely disagree to "my mind was on something other than the lesson," even though they were engaging in extraneous processing. Further research should be conducted on how to assess for distraction and extraneous processing, especially when participants are engaging in IVR learning.

Chapter IV: Experiment 3

Rationale and Hypotheses

The goal of Experiment 3 was to further investigate the findings from Experiment 2, specifically by trying to identify what may be serving as a source of distraction within an IVR lesson. As suggested by some of the comments from Experiment 2, IVR may be difficult to learn from, at least at first, because using the VR device is new to many learners. As such, a source of distraction may be due to the novelty of using the VR device for learning (Mayer et al., 2022). To investigate this question, I have used research on the pre-training principle (e.g., Mayer et al., 2002; Mayer & Fiorella, 2016, 2022) in an attempt to combat the potential novelty of using this device. The pre-training principle predicts that off-loading some learning to before a lesson can benefit learning outcomes by distributing the amount of information to be learned into two chunks. In this case, participants played a game in VR meant to familiarize participants with the interactive nature of the device prior to watching the lesson in order to off-load learning how to use the VR device. If the novelty of using IVR does serve as a distraction for learners, including pre-training focused on how to use the device should reduce the amount of distraction experienced in a lesson and thereby benefit learning outcomes.

The main question addressed in this research was: does allowing participants to become familiar with a VR device prior to showing a lesson in IVR benefit learning outcomes? More specifically, this research aimed to determine whether a VR pre-training experience benefited learning from a subsequent IVR lesson compared to no pre-training experience. A secondary question addressed in this research was: do executive function and working memory capacity relate to the effectiveness of a VR pre-training experience. I investigated these questions with a one-way between-subjects design in which participants either played a game in VR then watched the IVR lesson or just watched the IVR lesson then completed a posttest and completed working memory tasks.

The first hypothesis predicted that learning with IVR should be better when learners have time to experience VR technology prior to the lesson rather than not having prior experience with the device before the lesson. More specifically, it was predicted that those who played a game in IVR prior to learning with an IVR lesson (i.e., the IVR+ condition) would give lower ratings of distraction (both on a direct question assessing distraction and extraneous processing questions) during learning the lesson than those who only saw the IVR lesson (IVR condition; hypothesis 1a) and thus the IVR+ condition would perform better on a posttest than the IVR condition (hypothesis 1b). The second hypothesis predicted that learners who were better at managing incoming information should learn more information from a lesson, and this would only be true if distractions are present in the lesson. More specifically, it was predicted that posttest scores would correlate with scores on executive function tasks in the IVR condition (hypothesis 2a), but there would not be a correlation for the IVR+ group (hypothesis 2b). Finally, the third hypothesis predicted that holding capacity should impact learning only when distractions are present. More specifically, it was predicted that posttest scores would correlate with scores on working memory capacity tasks when learners were in the IVR condition (hypothesis 3a), but not in the IVR+ condition (hypothesis 3b).

Method

Participants and Design

The participants were 204 participants recruited from the Psychology Subject Pool at the University of California, Santa Barbara. However, 25 participants had moderate to high amounts of previous experience with IVR and were not included in the analyses. Additionally, 25 participants did not complete the study, so their data were not included in any analyses. This left 154 participants with complete data and low or no prior experience with VR. The mean age of the remaining participants was 19.32 years (*SD* = 1.37) and most of the participants were in their 1st or 2nd year at the university (*M* = 2.08 years, *SD* = 1.06). Additionally, the prior knowledge of the participants was somewhat low (*M* = 2.06, *SD* = .95), on a 5-point Likert scale from "1 – Very Low" to "5 – Very High." Of the sample, 112 participants identified as women, and 42 participants identified as men.

In a between-subjects design, there were 78 participants randomly assigned to the IVR lesson after pre-training on a VR game (IVR+ condition) and 76 participants randomly assigned to the IVR lesson without pre-training on a VR game (IVR condition). For the first set of analyses involving *t*-tests, the independent variable was the condition participants were in and the dependent variables were posttest performance, self-reported distraction rating, and subjective extraneous processing rating. For the second set of analyses involving correlation and regression, the predictor variables were executive function task scores, including scores on the Stroop task, n-back task, and flanker task, and working memory capacity task scores, including scores on the digit span task, Corsi block task, and operation span task. The outcome variable was posttest performance.

Materials

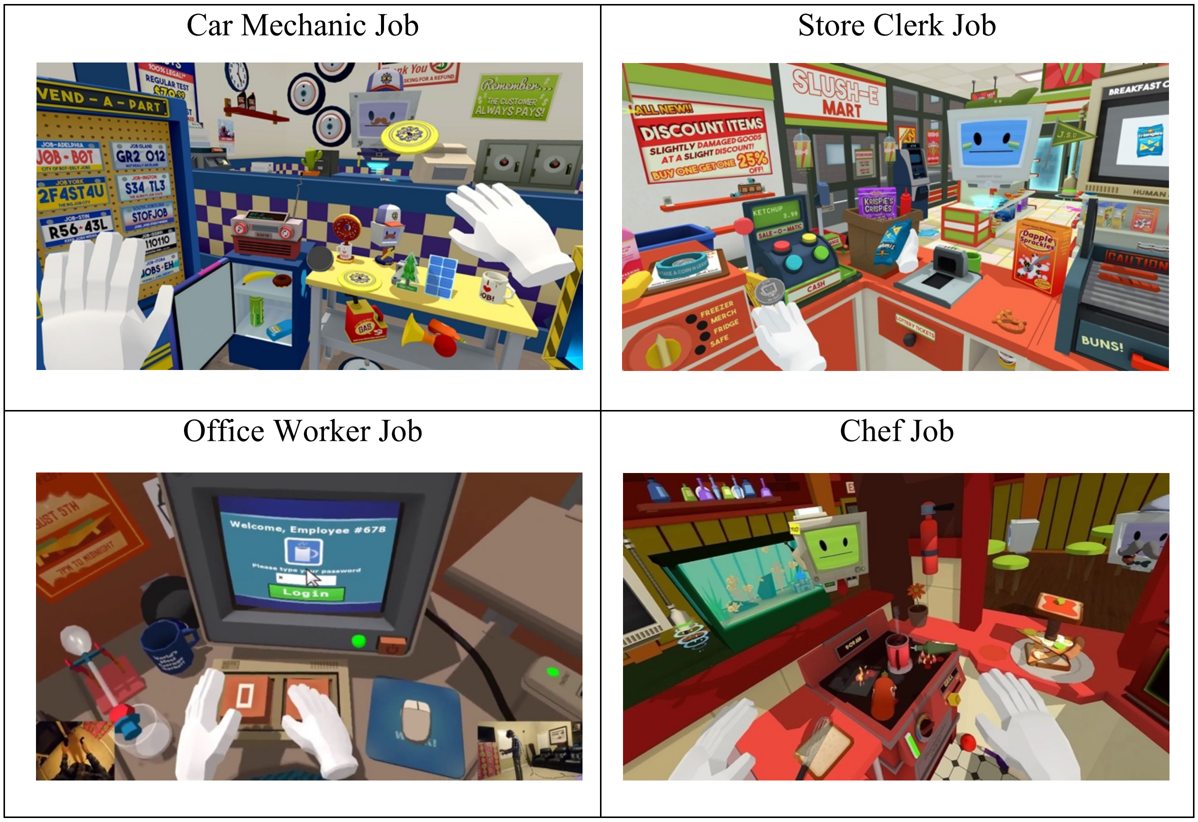
All the materials were computer-based or presented via a head-mounted display and were presented in a controlled laboratory setting. They included a prequestionnaire, pre-lesson VR experience, IVR lesson, posttest, postquestionnaire, and working memory tasks.

Prequestionnaire. The prequestionnaire was the same as Experiment 2. No participants reported being extremely prone to motion sickness, so everyone was able to participate in the VR experience.

Pre-Lesson VR Experience. For the IVR+ condition, participants were asked to play a game called "Job Simulator" in the VR headset for 10 minutes prior to moving to the IVR lesson. Job Simulator is a game created by Owlchemy Labs. The premise of the game is that computers have replaced humans in all jobs, and the VR game is a simulated experience that allows humans to experience what it is like to have a job. The participants were given the opportunity to choose one of four jobs: car mechanic, cook, office worker, or store clerk. See Figure 9 for images from these different VR experiences. Within each job, participants were given different objectives to complete that required them to look around the whole 360-degree space and use the hand controls to interact with objects in the environment. However, participants were not required to complete the objectives, as long as they interacted with the environment during the 10 minutes. During the game, participants were monitored and given prompts to engage with the environment if they were not interacting with the game. The rationale for the game was to give learners the experience of being in and interacting with immersive virtual reality. If the novelty of using a VR device is a large part of the reason that IVR is distracting to learn in, then giving learners experience

Figure 9.

Images from Different Job Simulator Experiences



with the device prior to learning should lessen how much learners need to focus on figuring out the device. When learners are not focused on how to use the device, which is irrelevant to the goal of the lesson, learners should be less distracted than if they do not get prior experience in VR.

Lesson. The lesson was the same IVR lesson as the one in Experiment 2.

Posttest. The posttest was the same as the one in Experiment 2 and was administered in the same way.

Postquestionnaire. The postquestionnaire was similar to the one in Experiment 2. The questions about the participants' experience with the lesson, subjective experience with cognitive load, presence, and cybersickness were the same from Experiment 2. McDonald's Omega was .78 for extraneous processing questions, .59 for essential processing questions, and .64 for generative processing questions. McDonald's Omega for the presence questions was .88. McDonald's Omega for the oculomotor questions was .60 and for the disorientation questions was .61. There were several additional questions added from the Experiment 2 version. First, several questions were added to assess the different sources of distraction in an IVR lesson, such as "The instructions for the lesson I saw distracted me from the lesson." "Interacting with the environment in the lesson distracted me from the content." and "Being immersed in the environment in the lesson distracted me from the content." all of which were answered on a 5-point Likert scale from "1 – Strongly Disagree" to "5 – Strongly Agree."

Lastly, for those who were in the IVR+ condition, there were several questions that asked participants to rate their experience with Job Simulator, including "I enjoyed playing Job Simulator." and "Playing Job Simulator distracted me from learning the material presented in the lesson." all of which were answered on a 5-point Likert scale from "1 – Strongly Disagree" to "5 – Strongly Agree." All the questions presented in the postquestionnaire are presented in Appendix C.

Working Memory Tasks. The same battery of working memory tasks from Experiment 2 was used to assess participants' ability to manage incoming information (executive function) and ability to hold information (working memory capacity).

Apparatus

For the lesson and VR pre-training experience, participants used the HTC Vive head-mounted display virtual reality system connected to a Dell Alienware computer. The lesson and game were operated through Steam software on the Dell computer. In addition to the head-mounted display, participants had wireless hand controls that they used to interact with the environment and receive haptic feedback from the virtual environment. The all the non-lesson material was presented on a Dell computer with a 20-inch color monitor.

Procedure

Participants started the study by signing the consent form. Once they agreed to participate, they completed the prequestionnaire, the digit span task, the Stroop task, and the Corsi block task. Then, they were randomly assigned to either the IVR+ condition or the IVR condition. If they were randomly assigned to the IVR+ condition, participants moved into the Job Simulator game and played for 10 minutes, then moved directly into the IVR lesson. Participants in the IVR condition started immediately with the IVR lesson, without playing the game. When they had completed the lesson, participants were thanked for their participation and told to return for the second part of the study a week later.

In the second session, participants started by completing the posttest. Once done with the posttest, the survey moved directly to the postquestionnaire questions. Then, participants were instructed to complete the n-back task, flanker task, and the operation span task. Once done with that, participants were thanked again for their participation and given credit for their participation. Each session of the experiment took no longer than an hour to complete. The study was approved by the university's Institutional Review Board (IRB).

Results

Posttest Scoring

The posttest was scored in the same way and with the same rubric as Experiment 2. Two independent scorers read each response and gave the participants scores. The correlation between the two scorers was strong, $r = .90$, $p < .001$. All differences in scoring between the scorers was discussed until 100% agreement. The reliability of the posttest was determined using McDonald's Omega and was .76.

Outliers. In this data, outliers were determined to be participant scores on the measures of working memory in which the z-score exceeded positive or negative three. In this data set, four participants were removed from the data because of their outlier scores, including one participant due to their low n-back task scores, and three participants due to their high flanker task scores. Removing the outliers did not change the pattern results.

Are the Groups Equivalent on Key Characteristics?

Before proceeding with further analyses, I checked that random assignment was successful by comparing the IVR+ and IVR conditions. Means and standard deviations on basic characteristics for the two groups are presented in Table 11. There was not a significant difference between groups based on age, $t(146) = -.14$, $p = .887$; class level, $t(148) = .06$, $p = .951$; subjective prior knowledge, $t(148) = -.93$, $p = .337$; or gender ratios, χ^2 $(1, N = 150) = 1.91$, $p = .167$. As such, it was determined that random assignment was successful in creating equivalent groups for these key characteristics.

Table 11

Means and Standard Deviations for Experiment 3

	IVR+ Condition		IVR Condition	
	M	*SD*	*M*	*SD*
Age	19.33	1.49	19.36	1.25
Class Level	2.11	1.11	2.09	1.01
Prior Knowledge	1.99	.95	2.14	.94
Posttest Score	12.66	4.73	12.45	4.58
Presence	3.97	.80	4.16	.61
Enjoy	4.51	.53	4.70	.54
Interesting	4.33	.68	4.43	.78
More Lessons	4.30	.75	4.49	.78
Effective	4.11	.86	4.16	.91
Motivated	4.08	.86	4.35	.80
Distraction from Lesson	1.83	1.04	1.70	.96
Extraneous Processing	2.07	.78	1.94	.78
Distracting Instructions	1.66	.83	1.73	.95
Distracting Interactions	2.08	1.20	2.08	1.18
Distracting Environment	1.95	1.14	1.79	1.13

Hypothesis 1: Do Learners Who Are Given Prior Experience with VR Report Less Distractions and Have Better Learning than Learners Who Are Not?

Hypothesis 1 predicted that the IVR+ condition should experience less distraction and thus report less distraction and less extraneous processing compared to the IVR condition. Additionally, hypothesis 1 predicted that the IVR+ condition should have better learning due to the reduced distraction compared to the IVR condition. To test this hypothesis, posttest scores, responses to the questions about the distractions that participants felt during learning, and responses to the cognitive load questions focused on extraneous processing were analyzed using *t*-tests. Means and standard deviations on these variables for the two groups are reported in Table 11. First, there was not a significant difference between the IVR+ and IVR conditions on how much participants felt distracted during the lesson, $t(148) = .77$, $p = .441$. See Figure 10a for a bar graph of distraction across groups. Additionally, there was not a significant difference between these groups on how distracting the instructions were, $t(147) = -.47$, $p = .640$ (Figure 10b), how distracting it was to interact with the environment, $t(147) = .82$, $p = .413$ (Figure 10c), nor how distracting it was to be immersed in the environment, $t(147) = -.02$, $p = .987$ (Figure 10d). For the amount of extraneous processing participants reported feeling, there was not a significant difference between the groups, $t(148) = .77$, $p = .331$ (Figure 10e). These findings do not support hypothesis 1a.

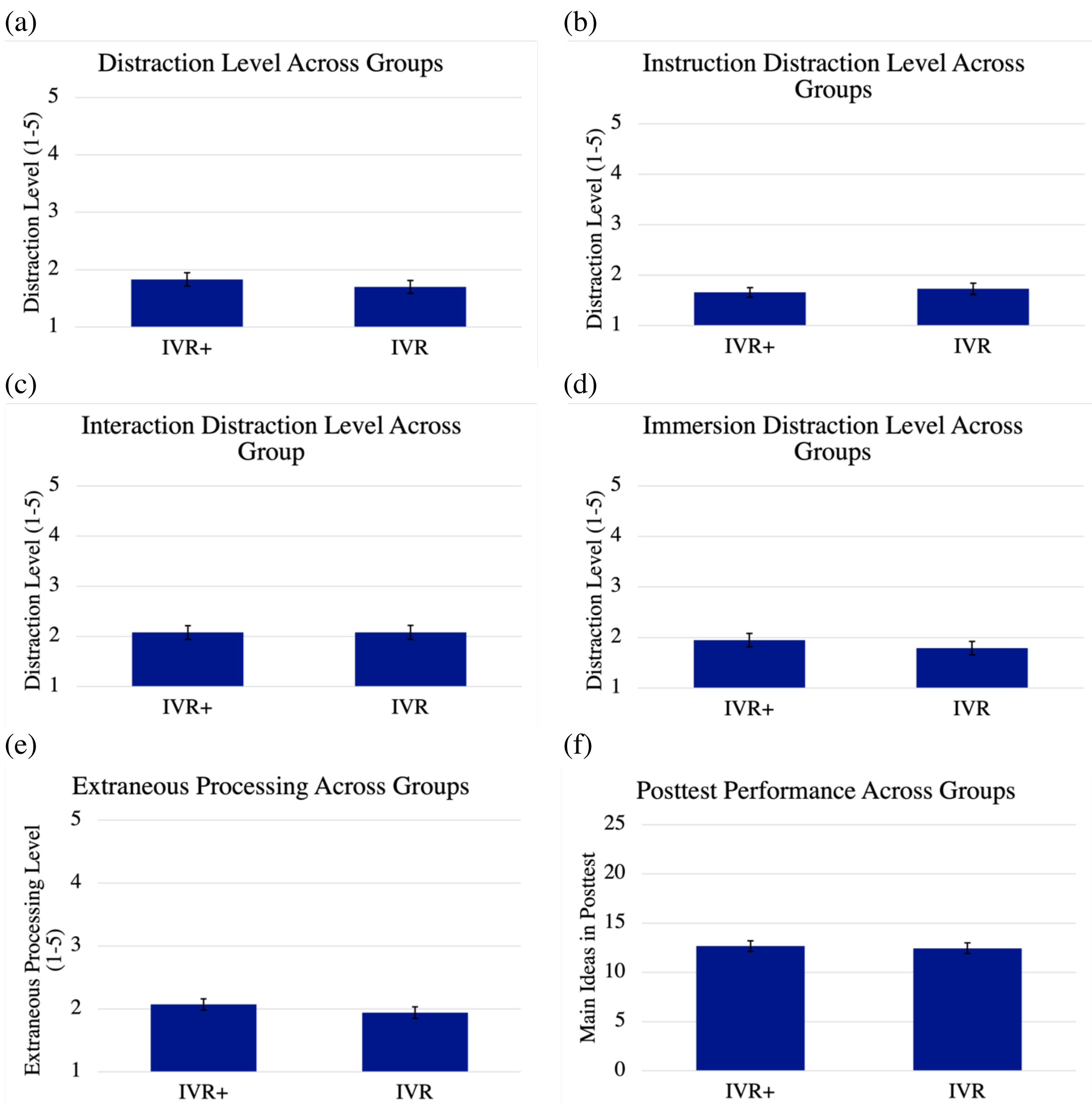
For posttest scores, there was also not a significant difference between the groups, $t(148) = .28$, $p = .781$, not supporting hypothesis 1b. See Figure 10f for a bar graph of distraction across groups. These findings do not support hypothesis 1 and indicate that playing a game prior to learning a lesson in IVR did not reduce the number of distractions that learners engaged with and did not benefit learning compared to those who did not play the game prior.

Hypothesis 2: Do Posttest Scores Correlate with Scores on Executive Function Tasks when Learning With or Without Pre-Training in VR?

Hypothesis 2 predicted that posttest scores should correlate with executive function task scores for those in the IVR condition because they should have had more distractions to deal with during learning (hypothesis 2a). Hypothesis 2 also predicted that posttest scores should not correlate with executive function task scores for those in the IVR+ condition because the distractions should have been reduced by the pre-training (hypothesis 2b). The

Figure 10.

Mean Comparisons for Distraction Levels and Posttest Performance for Experiment 3



(a) subjective experience of general distraction across IVR+ and IVR conditions (b) subjective experience of distraction from the instructions across IVR+ and IVR conditions (c) subjective experience of distraction from the interacting across IVR+ and IVR conditions (d) subjective experience of distraction from the immersion across IVR+ and IVR conditions (e) subjective experience of extraneous processing across IVR+ and IVR conditions (f) posttest scores across IVR+ and IVR conditions main analysis to answer this hypothesis was to correlate posttest scores with scores on executive function tasks for each condition. Thus, the data first was split into the two separate conditions. For each condition, a correlation matrix was constructed to determine the relationships between posttest scores and the three executive function tasks, which is presented in Table 12.

For the IVR condition, n-back scores were significantly correlated with posttest scores, $r = .30$, $p = .010$, partially consistent with hypothesis 2a. See Figure 11a and 11c for scatterplot of posttest scores with n-back task score and Stroop task score, respectively. None of the other executive function tasks was significantly correlated with posttest scores for the IVR condition. For the IVR+ condition, the Stroop task scores, $r = -.29$, $p = .011$, and n-back task scores, $r = .28$, $p = .015$, were significantly correlated with posttest scores, partially inconsistent with hypothesis 2b. See Figure 11b and 11d for scatterplot of posttest scores with n-back scores and flanker scores, respectively. Flanker task was the only executive function task that was not significantly correlated with posttest scores for the IVR+ condition. However, in both the conditions, flanker task did have a small to medium correlation, but it was not significant.

To supplement these analyses, a stepwise regression was used for each condition to determine if there was a predictive relationship between the executive function tasks and posttest performance. The three predictors were included in the regression and the program selected only those that provided predictive power in the final output. For the IVR condition, only the n-back task had predictive power for the posttest scores, so it was the only variable of the three that was included in the stepwise regression, $R = .30$, $R^2 = .09$, $F(1, 71) = 6.92$, p

Table 12.

Correlation Matrix for IVR and IVR+ Lessons in Experiment 3

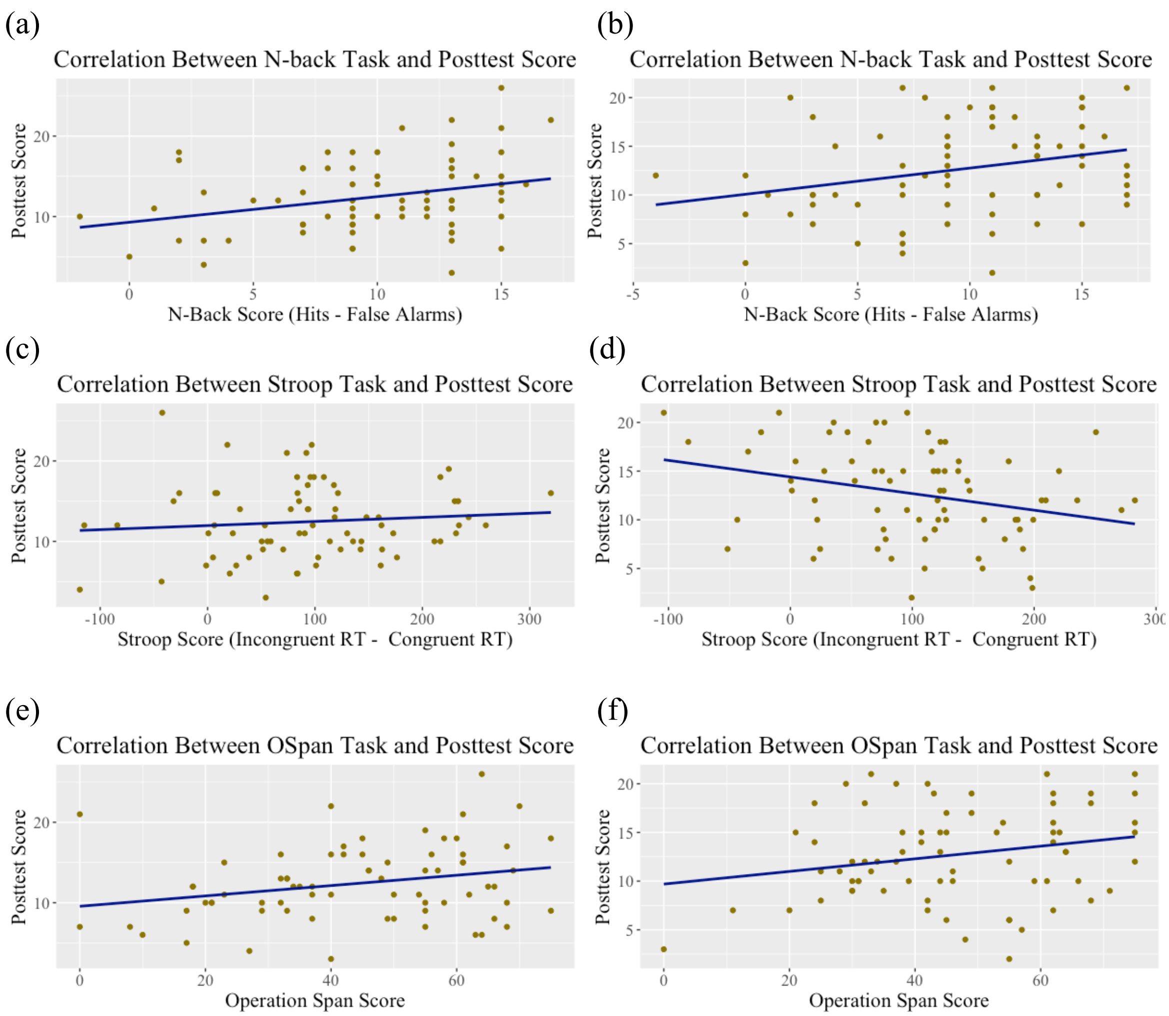
	Posttest	Stroop	N-back	Flanker	Digit Span	Operation Span	Corsi
IVR+ Lesson							
Posttest	1.00	-.29*	.28*	-.15	-.02	.23*	-.08
Stroop		1.00	-.07	-.05	.03	-.10	-.08
N-back			1.00	.08	.03	.27*	-.23*
Flanker				1.00	.03	-.11	-.19
Digit Span					1.00	.40***	.18
Operation Span						1.00	.12
Corsi							1.00
IVR Lesson							
Posttest	1.00	.10	.30**	.22	.15	.26*	-.08
Stroop		1.00	.16	-.18	-.02	.03	.09
N-back			1.00	.02	.12	.34**	.11
Flanker				1.00	.09	.17	.12
Digit Span					1.00	.39***	.21
Operation Span						1.00	.07
Corsi							1.00

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

= .010. The n-back task was related to posttest performance in the predicted direction, such that as n-back task score increased–indicating an increase in executive function–posttest scores increased, $B = .32$, $\beta = .30$, $t = 2.63$, $p = .010$. For the IVR+ condition, Stroop task and n-back task had predictive power for the posttest scores, $R = .39$, $R^2 = .15$, $F(1, 72) = 6.42$, $p = .003$. N-back task was a stronger predictor for posttest score and was related to posttest performance in the predicted direction, such that as n-back task scores increased–indicating an increase in executive function–posttest scores increased, $B = .25$, $\beta = .26$, $t = 2.41$, $p = .018$. Stroop task was related to posttest performance in the predicted direction as well, such that as Stroop task scorse decreased–indicating an increase in executive function–

Figure 11.

Correlation Between N-back & Stroop and Posttest Scores in Experiment 3



(a) Correlation between n-back task and posttest performance for the IVR condition (b) correlation between n-back task and posttest performance for IVR+ condition (c) correlation between Stroop task and posttest performance for IVR condition (d) correlation between Stroop task and posttest performance for IVR+ condition (e) correlation between operation span task and posttest performance for IVR condition (f) correlation between operation span task and posttest performance for IVR+ condition

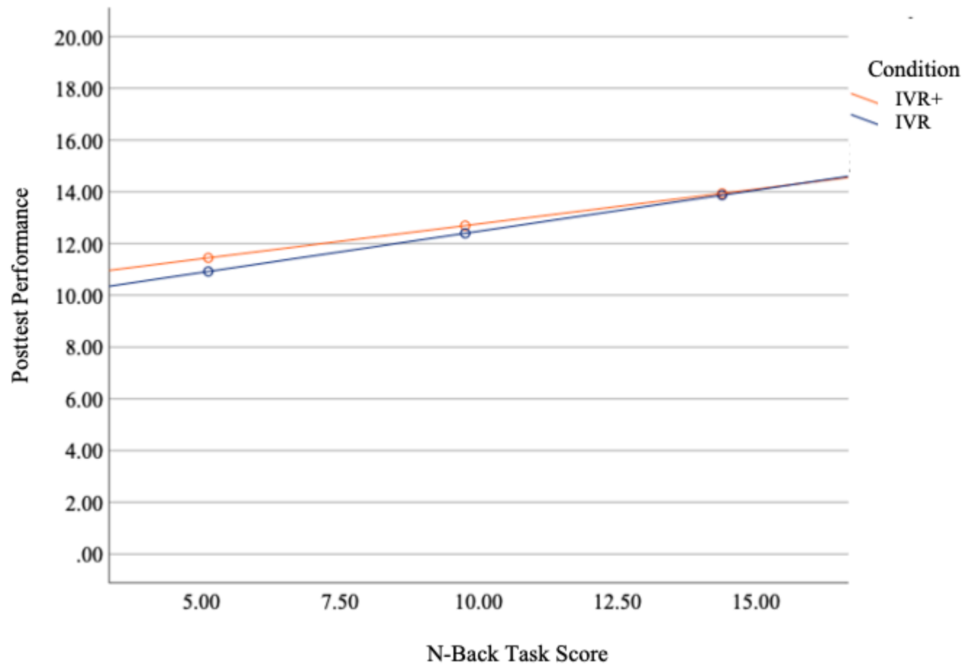
posttest scores increased, $B = -.02$, $\beta = -.27$, $t = -2.49$, $p = .015$. Overall, there is some evidence that executive function is related to learning outcomes when students are learning with IVR, regardless of whether they receive pre-training in VR prior to learning.

As a secondary analysis to compare the relationship between n-back task scores and posttest performance across the two conditions, a moderation analysis was conducted. In this analysis, n-back task score was the predictor variable, posttest performance was the outcome variable, and lesson type was the moderator. See Figure 12 for a graph of the moderation. The moderation analysis demonstrated that there was not a significant interaction between n-back score and lesson type of learning, $p = .762$. This is not surprising as the correlations between n-back task scores and posttest performance were similar across the two conditions.

A moderation analysis was also conducted to compare the relationship between Stroop task scores and posttest performance across the two conditions. In this analysis, Stroop score was the predictor variable, posttest performance was the outcome variable, and lesson type was the moderator. See Figure 13 for a graph of the moderation. The moderation analysis demonstrated that there was a significant interaction between Stroop task score and lesson type on learning, $p = .015$. This suggests that the relationship between Stroop task score and posttest performance was different for the two conditions. Investigating further, the moderation analysis revealed that for the IVR+ condition, the slope of the line was significantly different from zero, $t = -2.60$, $p = .010$, such that an increase in Stroop task score—or a decrease in executive function—was related to a decrease in posttest performance. In the IVR condition, the slope of the line was not significantly different from zero, $t = .83$, $p = .406$. This suggests that when participants were given training in IVR prior to the lesson, their ability to manage incoming information during the lesson seemed to play an important role in helping them learn but this was not the case for participants who were not given training.

Figure 12.

Moderation Analysis for N-Back Task Score on Posttest Performance By Condition

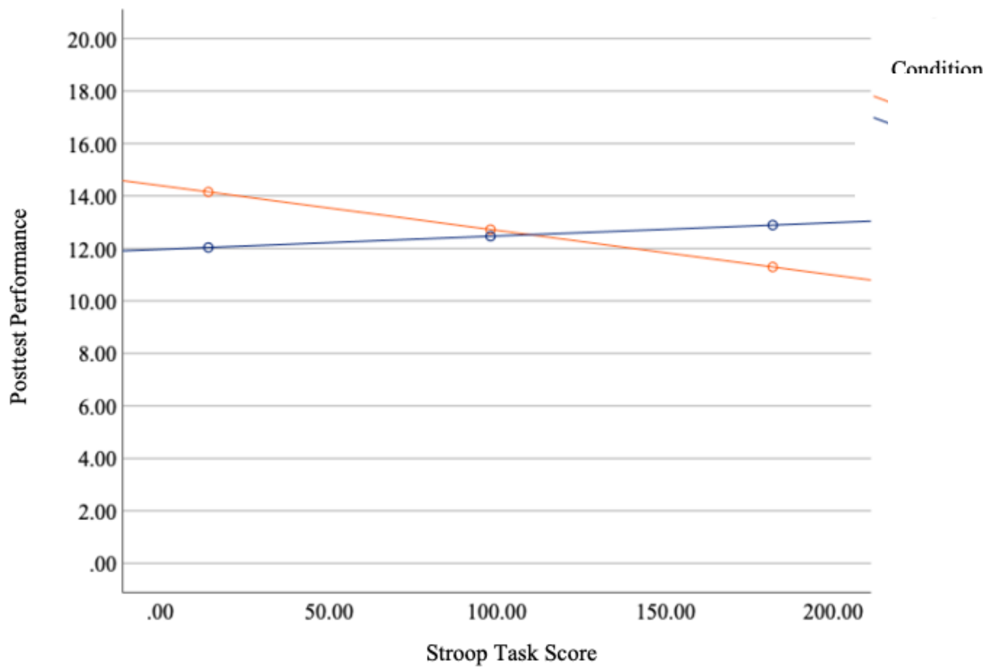


Hypothesis 3: Do Posttest Scores Not Correlate with Scores on Working Memory Capacity Tasks when Learning With or Without Pre-Training in VR?

Hypothesis 3 predicted that posttest scores should correlate with working memory capacity task scores for the IVR condition (hypothesis 3a), but not for the IVR+ condition because those who have training in VR prior to the lesson should have fewer distractions to deal with (hypothesis 3b). The main analysis to answer this hypothesis was to correlate posttest scores with scores on working memory capacity tasks. A correlation matrix was created with the posttest scores and working memory capacity tasks scores, displayed in Table 12. For the IVR condition, operation span task scores were significantly correlated

Figure 13.

Moderation Analysis for Stroop Task Score on Posttest Performance By Condition



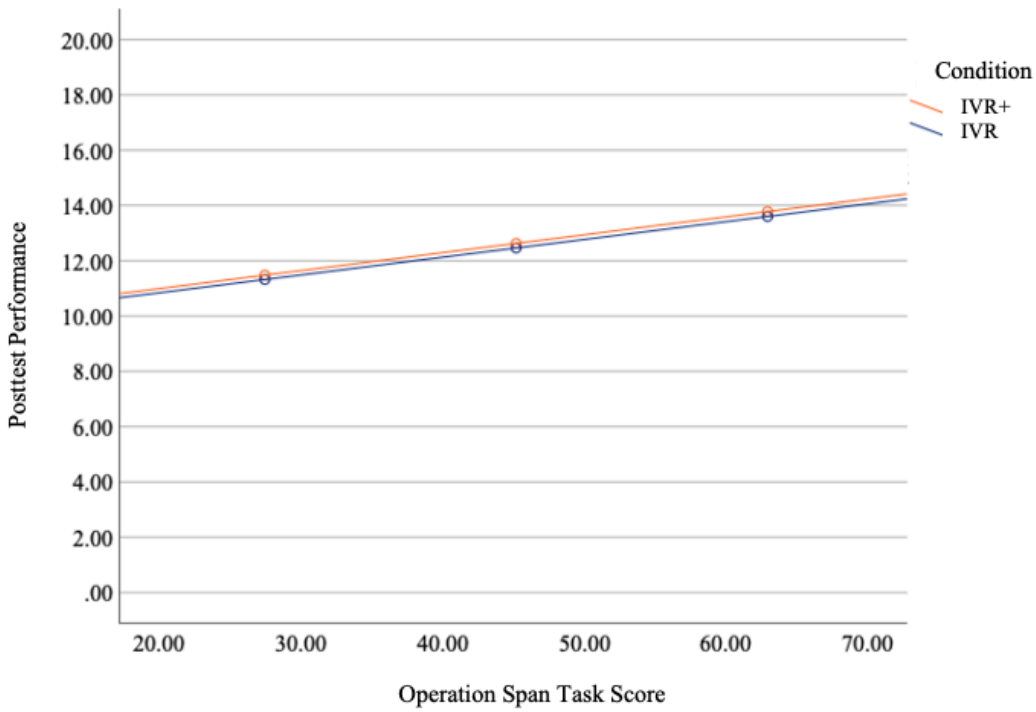
with posttest scores, $r = .26$, $p = .029$, partially consistent with hypothesis 3. See Figure 11e for scatterplot of this relationship. None of the other working memory capacity tasks was significantly correlated with posttest scores for the IVR condition. For the IVR+ condition, once again operation span task scores were significantly correlated with posttest scores, $r = .23$, $p = .049$, partially consistent with hypothesis 3. See Figure 11f for scatterplot of this relationship. None of the other working memory capacity tasks was significantly correlated with posttest scores for the IVR+ condition.

To supplement these analyses, a stepwise regression was used for each condition to determine if there was a predictive relationship between the working memory capacity tasks and posttest performance. The three predictors were included in the regression and the program selected only those that provided predictive power in the final output. For the IVR condition, only the operation span task had predictive power for the posttest scores, so it was the only variable of the three that was included in the stepwise regression, $R = .26$, $R^2 = .07$, $F(1, 69) = 5.00$, $p = .029$. As operation span task score increased–indicating a larger working memory capacity–posttest scores increased, $B = .06$, $\beta = .26$, $t = 2.24$, $p = .029$. For the IVR+ condition, operation span task had predictive power for the posttest scores, $R = .23$, $R^2 = .05$, $F(1, 71) = 4.00$, $p = .049$. Once again, as operation span task score increased–indicating a larger working memory capacity–posttest scores increased, $B = .07$, $\beta = .23$, $t = 2.00$, $p = .049$. Overall, there is some evidence that working memory capacity is related to learning outcomes when students are learning with IVR, regardless of whether they receive pre-training in VR prior to learning.

As a secondary analysis to compare the relationship between operation span task scores and posttest performance across the two conditions, a moderation analysis was conducted. In this analysis, operation task score was the predictor variable, posttest performance was the outcome variable, and lesson type was the moderator. See Figure 14 for a graph of the moderation. The moderation analysis demonstrated that there was not a significant interaction between operation span task score and lesson type of learning, $p = .762$. This is not surprising as the correlations between operation span task scores and posttest performance were similar across the two conditions.

Figure 14.

Moderation Analysis for Operation Span Task Score on Posttest Performance By Condition



Exploratory Analyses

To expand on some of previous findings about the role that training in the VR device can play in learning using an IVR lesson, several other analyses were conducted using the data from the 25 participants with moderate to high amounts of prior VR experience that were not included in the original data. Participants were split into two groups, those with no-to-low prior experience in VR and those with moderate-to-high prior experience in VR. There was not a significant difference in posttest performance between the lower prior experience group (M = 12.41, SD = 4.71) and the higher prior experience group (M = 13.20, SD = 5.67), $t(177) = -.76$, p = .451. Additionally, to make this even more constrained, participants with low prior experience in VR were removed to compare only those who had no prior experience in VR to those with moderate to high amounts of experience. Once again, there was not a significant difference in posttest performance between the no prior experience group (M = 12.58, SD = .53) and the higher prior experience group, $t(103) = -.55$, $p = .584$. These analyses help demonstrate that the impactful distractions that participants experience in a lesson are likely coming from a source other than figuring out how to use the VR device. However, this is not conclusive evidence as the sample of participants with moderate to high prior experience with VR is small.

Several other analyses were of interest to this research, but not specifically predicted in the hypotheses. Means and standard deviations on the ratings of the participants' perceptions of the lesson are provided in Table 11. Sense of presence was compared between the IVR condition and the IVR+ condition, as having knowledge of how to use the VR device may lend itself to experiencing more presence while learning. However, this was not true as there was not a significant difference between the two conditions, $t(148) = -1.58$, $p = .117$.

There were also some concerns that those who were able to play a fun game right before learning a more serious lesson (IVR+ condition) may not have the same experience as those who learn only in IVR without pre-training (IVR condition). As such, ratings of enjoyment, interest, desire to learn similarly in the future, subjective effectiveness, and motivation were compared between the two conditions. There was a significant difference between the two conditions on enjoyment, $t(148) = -2.17$, $p = .032$. Those in the regular IVR lesson reported that the lesson was more enjoyable than the IVR+ lesson. Additionally, there was a significant difference between the two conditions on motivation, $t(148) = -2.01$, $p = .047$. Those in the regular IVR lesson reported that the lesson was more motivating than the IVR+ condition. Both of these findings may have occurred because playing a fun game in IVR right before learning a lesson that is less interactive than the game may have been a difficult transition. However, there was not a significant difference between the two conditions based on interest, $t(148) = -.87$, $p = .387$; desire to learn similarly in the future, $t(148) = -1.47$, $p = .143$; and subjective effectiveness, $t(148) = -.40$, $p = .694$. As such, although there were some differences in experience between the two conditions, the two experiences were not completely different.

I was also interested in understanding how beneficial participants who were in the IVR+ condition felt about the effectiveness of the pre-training experience. Participants were asked if they enjoyed playing Job Simulator, if they thought it was helpful for learning about the VR device, if they thought the experience was a waste of time, and if they were distracted by playing the game all on a 5-point Likert scale from "1 – Strongly Disagree" to "5 – Strongly Agree." Participants reported that they generally enjoyed playing Job Simulator ($M = 4.42$, $SD = .86$) and felt it was a good way to learn how to use the VR device ($M = 4.21$, $SD = 1.03$). Additionally, they generally did not think of playing Job Simulator as a waste of time ($M = 2.21$, $SD = 1.20$) nor did they find it distracting to learning the lesson ($M = 1.87$, $SD = 1.13$).

Comments on Lesson. Open-ended comments about the experience were encouraged, and as such, participants' comments are briefly analyzed here. To begin, there were a similar number of comments submitted for both the IVR+ condition (28 comments) and the IVR condition (27 comments). Across the two conditions, the nature of the comments was similar, focusing mainly on comments about how the participants felt about the experience and the distractions experienced during the lesson.

In both conditions, many participants made comments that indicated their positive experience with the lesson (15 comments for the IVR+ condition – 53.57% -- and 17 comments for the IVR condition – 62.96%). For the participants in the IVR+ condition, there were comments, like "I really enjoyed the lesson and learned more about the acidification of oceans," "I thought that the VR lesson was a very exciting way to learn about new material," and "I enjoyed that it was easy and fun. A good way to motivate and start the actual learning process." Participants in the IVR condition had similar comments, like "I thought it was super fun and I also think having people do the immersive VR could be a very very persuasive and effective way of teaching people, as it really stuck with me and felt powerful," "It was great," and "The simulation was amazing."

Despite having positive views on the experience, there were numerous comments that mentioned the distractions that participants faced in the lesson (11 comments for the IVR+ condition – 39.29% -- and 6 comments for the IVR condition – 23.08%). For the participants in the IVR+ condition, there were comments, like "although I do think I learned from this, I was distracted with the sea life and looking around which made me forget informatrion that could have helped me today," "The novelty of using VR was somewhat distracting to the experience of learning about ocean acidification," and "The noise was a little too much." Participants in the IVR condition had similar comments, like "I liked the VR, but I think that it would have been a little better if I was simply supposed to watch the lesson, rather than actively participate in it, such as having to look for shells," "I wish there were captions as the narrator was speaking," and "it was my first time using VR, so it was all very exciting, and I would find it difficult to focus on the lesson. But once I was used to it, it was okay."

The themes to the comments are similar to the themes that came up in the IVR condition in Experiment 2, indicating that although participants do enjoy using IVR for learning, they do recognize that it can be difficult to learn with the present distractions. This demonstrates that, despite some of the participants having pre-training in IVR (the IVR+ condition), many participants regardless of their condition mentioned experiencing distraction while learning in IVR. As such, it seems as though the pre-training was not particularly effective in reducing the distraction of learning in IVR, consistent with the quantitative findings. However, one interesting finding is that people in the IVR+ condition seemed to be more aware of the fact that learning in IVR could be distracting, determined by the difference in the number of comments made. This may suggest that the pre-training was beneficial in having students notice distraction, but not necessarily helpful in mitigating the impact of the distractions.

Discussion

The main findings of Experiment 3 demonstrated that having participants engage in an IVR game prior to learning in IVR does not mitigate the distractions that are brought on by learning with this technology. This was demonstrated through there being no differences between the two conditions in reported distractions and posttest performance. Additionally, in both conditions, executive function task scores related to posttest scores, indicating that some level of distraction was experienced in both lessons. This indicates one of two things: (1) the pre-training experience was not enough to actually allow learners to develop an understanding of how to use the VR device or get used to the IVR experience, or (2) something other than the novelty of the VR experience is creating distractions to a lesson taught in IVR.

One interesting result was that there was a moderating effect of lesson type on the relationship between Stroop task and posttest performance in which learners' ability to manage incoming information impacted learning only in the condition where participants had pre-training in VR. This goes against what would be expected if the novelty of the lesson is causing distraction for learners. However, this could suggest that having training in VR prior to the lesson allows participants to become more comfortable with learning in IVR and thus opens them up to engaging in more distraction. That is, if someone feels comfortable using IVR and knows how to make use of the 360 degrees of visual information, they may be more likely to move, spin, and interact with the lesson, which increases the number of distractions, or irrelevant details, they will experience. This is consistent with some of the comments in the qualitative portion of the experiment. More research needs to be done to fully make conclusions about this effect.

The open-ended comments also added to the evidence that participants did not differ much in their experiences with the lesson, regardless of if they had pre-training with VR or not. These comments demonstrated an overwhelmingly positive view of using IVR for learning, as it is an exciting and fun device to use. However, a handful of participants in both conditions made comments that focused on how they were distracted within the lesson, which may have hurt their learning.

In Experiment 3, working memory capacity, measured by operation span task, was shown to correlate with learning. The previous experiments have not found evidence for this, but in Experiment 3, across both conditions, better performance on the operation span task–indicating a larger working memory capacity–predicted better posttest scores. This finding may indicate that previous literature discussing the distinction between complex and simple span tasks (e.g., Daneman & Carpenter, 1980; Ilkowska & Engle, 2010; Redick et al., 2012) is important to consider in research on learning, as complex span tasks may be able to tell us something important about individual differences in learning from technology. Additionally, this finding may be due to the fact that complex span tasks require some executive function to successfully complete the task; in the operation span task, participants have to try to hold the letter sequence while also manipulating and working with the math equation, so they have to manage what information is being worked on at one time (Engle, 2002).

Chapter V: General Discussion

Empirical Contributions

The main goal of this set of experiments was to investigate which types of learning environments are effective *for whom*. Previous literature has mainly been focused on whether different instructional design techniques and learning environments are effective for learning but have not focused much on how individual differences in learners could be playing a role in how people learn. This research helps expand previous research by narrowing in on individual difference factors, specifically the ability to manage and hold incoming information, and how these factors can have impactful effects on how people learn in the same learning environments.

From this series of experiments, one of the main, consistent findings was that performance on classic measures of executive function was related to posttest scores when learning from lessons that had distracting features. In Experiment 1, this was demonstrated by the relationship between tasks measuring executive function and posttest performance when learning from multimedia lessons with distracting features, but not in multimedia lessons without distracting features. In Experiment 2, this was demonstrated by the relationship between tasks measuring executive function and posttest performance when learning from an IVR lesson, but not from a video lesson. In Experiment 3, this was demonstrated by the relationship between tasks measuring executive function and posttest performance in both conditions, as both conditions presented lessons in IVR.

Across all three experiments, learners who were better at managing incoming information while learning (as measured by classic tasks of executive function) performed better on posttests when learning from a lesson that had distracting features compared to learners who were worse at managing incoming information. This contrasts with the finding that a learner's ability to manage incoming information did not relate to learning from lessons without distracting features, such as the lesson presented in Study 1c or the slideshow lesson in Experiment 2. Within these lessons, one's ability to manage incoming information did not relate to learning.

These findings make logical sense; lessons that have distracting features require participants to use their ability to manage incoming information to sort the relevant information from the irrelevant information and ignore the irrelevant information. However, because individuals differ in their ability to manage incoming information, there are various impacts of distracting environments on different learners. If a lesson has distracting information, learners who are able to ignore the irrelevant information (better at managing incoming information) will be able to fill their limited working memory with only relevant information from the lesson while learners who are unable to ignore the irrelevant information (worse at managing incoming information) will fill their limited working memory with both relevant and irrelevant information and hit the limit sooner. As such, those who are better at managing incoming information will have better learning outcomes than those who are worse at managing incoming information.

Another important finding from this set of studies is that working memory capacity did not consistently significantly relate to posttest performance for the lessons. In Experiments 1 and 2, working memory capacity did not significantly correlate with posttest scores for both the distracting lessons and the non-distracting lessons. This would indicate that the amount of information that someone can hold at one time does not seem to play an important role in how much information they learn. As such, it would seem that the ability to manage incoming information is more important than the ability to hold new information to consider in learning from distracting lessons.

However, this pattern of results changed in Experiment 3; working memory capacity, specifically measured by operation span task, did correlate with posttest scores such that as one's capacity was larger, the more they were able to learn. In terms of statistical significance, these findings conflict with the findings from Experiment 2. However, investigating the correlation values, operation span task, in both Experiment 2 and 3, had consistently small to moderate correlations with posttest scores, suggesting that working memory capacity may be playing some role in learning in all learning environments, distracting and not distracting. If the ability to hold new information does play an important role in learning, individual differences on complex span tasks, like operation span task, seem to be more relevant to learning than individual differences on simple span tasks, like digit span task and Corsi block task.

One reason that operation span task may have differed from the simple span tasks is that operation span task requires a degree of executive function during the task (Engle, 2002). As such, it may be the case that this task is picking up the importance of executive function in learning and presenting correlations between the task and posttest learning. This would lead to the conclusion that holding capacity itself does not seem to play a role in learning.

Through this research, it was also found that people had similar posttest scores when learning from both the IVR and slideshow conditions, specifically in Experiment 2, but ability to manage incoming information only seemed to predict posttest outcomes when learning from IVR. On the surface level, this finding seems to contradict the coherence principle, but with the addition of the reliance on individuals' ability to manage incoming information, it seems as though the coherence principle is important to consider for certain learners. If just comparing the means of the groups, the coherence principle, stating that irrelevant information should be removed to benefit learning, does not seem to be supported by this research. But, with the individual difference component in mind, the coherence principle seems to be important to consider for learners who have a harder time managing incoming information. As such, this research helps demonstrate an important boundary condition for when following the coherence principle is most important.

Furthermore, this finding may help explain some of the conflicting findings in previous research on the benefits of learning in IVR compared to more traditional learning environments. In previous literature on the use of IVR in learning environments, there has been a lot of inconsistent findings, with some studies finding that IVR is better than more traditional learning mediums, some finding they are the same, and some finding IVR is worse. However, across these previous studies, individual differences, especially those regarding managing incoming information, have not been taken into account before now.

The differences that are seen in this previous literature may be explained by the lack of control over individual differences, as this set of experiments has demonstrated that an individual's ability to manage incoming information has important implications in their learning from an IVR lesson. If a sample of participants has a stronger ability to manage incoming information, their learning is likely not going to be negatively impacted by the distractions in the lesson and thus they are able to benefit more from the increased motivational and affective benefits of the IVR lesson (Flavia Di Natale et al., 2018; Lee et al., 2017; Makransky et al., 2019b; Makransky & Petersen, 2021; Parong & Mayer, 2018). However, if a sample of participants has weaker ability to manage incoming information, they are likely going to struggle with the distractions and not be able to learn as much in an IVR lesson compared to a more traditional lesson. As such, this research has demonstrated that making conclusions about the benefits of learning environments is not as simple as 'yes, IVR is helpful' or 'no, IVR is not helpful.' Instead, the answer must take into account specifics about the learner to determine which learning environment is going to be conducive to learning.

Lastly, this research has demonstrated, specifically in Experiment 3, that mitigating the distracting features of an IVR lesson is more difficult than briefly introducing learners to how to use a VR device. Even with a 10-minute introduction to a highly interactive game in IVR, learners did not do better on a posttest compared to those who did not receive this pre-training. Additionally, learners' ability to manage incoming information related to learners' posttest scores, suggesting that even with pre-training, learners still needed to use their ability to manage incoming information to ignore distractions and irrelevant information when learning in IVR. This finding suggests that the type of distractions learners face within an IVR lesson either needs more than 10-minutes of experience to mitigate or that the distractions are more inherent to the design of the lesson.

The results of Experiment 3 also demonstrated that, at least in terms of the Stroop task, training in VR prior to a lesson may contribute to the reliance on one's ability to manage incoming information. For the IVR+ condition, but not the IVR condition, participants who were better able to manage incoming information performed better on the posttest than those who were worse at managing incoming information. This could suggest that with more familiarity and understanding of using VR technology, learners may be more comfortable using the device and have to deal with more distractions from increased interactivity and movement in the device.

Lastly, the exploratory analyses that focused on how learners with high amounts of prior VR experiences performed on the posttest compared to learners with low amounts of VR experience suggests that it is more likely the second option, as even with moderate to high amounts of experience with VR, learners still were not able to learn more in an IVR lesson. As such, more work needs to be done to identify which specific features of an IVR lesson increase distraction and how these distractions can more effectively be mitigated.

Theoretical Contributions

These findings help expand theories of learning, like cognitive load theory and the cognitive theory of multimedia learning. Cognitive load theory (Paas & Sweller, 2022; Sweller, 1994, 2020) and cognitive theory of multimedia learning (Mayer, 2021, 2022) explain that while learning, individuals' working memory can be impacted by different types of demands on processing capacity, and as such have different consequences on learning, depending on the type of demand. For lessons that have distracting features or use distracting technology, extraneous load (in CLT) or extraneous processing (in CTML) are important to consider. Both theories predict that for lessons that induce an increased amount of extraneous processing, learning outcomes are poorer. However, these theories do not discuss how individual differences in cognition, specifically ability to manage incoming information, can equip some learners for dealing with extraneous processing better than others. As such, lessons that have distracting information or use distracting technology do not create the same amount of extraneous processing for all learners.

Both of these theories explain that the work individuals are doing to learn new material is all occurring within working memory (Mayer, 2021, 2022; Paas & Sweller, 2022; Sweller, 1994, 2020). However, these theories do not account for the fact that people vary in the capabilities of their working memory. For these theories to help make the best predictions about when students will be able to learn best, this research demonstrates that it is important to take into account learners' individual differences in the ability to manage incoming information. This is especially true for lessons that put high demands on learners' working memories, such as those that present many irrelevant details to learners.

Practical Implications

From this research, several practical implications arise. First, when learning from a lesson, learners' ability to manage incoming information is an important predictor of how well they will be able to learn the material, if they are learning from a lesson that has distractions. This is true for more traditional learning environments that are poorly designed, like a multimedia video lesson, and learning environments that use technology, like immersive virtual reality. As an instructor, this is vital to consider because using this type of material in an instructional setting contributes to inequity in the classroom. As the push to incorporate IVR into classrooms and learning environments has increased drastically in the last few years, this research is essential in recognizing how this can add to the inequity in learning within a class. As such, these findings suggest that instructors be mindful about the material they use in their classes until more research can be done on what can effectively mitigate distractions in both IVR lessons and more traditional lessons with distracting features.

Second, the findings of this research show how necessary it is to continue to conduct research on the role that individual differences play in learning. Although some literature has investigated the role that prior knowledge has in learning (e.g., Chen & Huang, 2013; Fiorella & Mayer, 2016; Gilmore & Papadatou-Pastou, 2009; Minnaert, 1999), a large portion of prior research on learning has focused on general principles that benefit all learners. The trend of investigating how individual differences impact learning experiences needs to become more widely used as there are important consequences of learning environments on learners.

Lastly, this research demonstrates an important need for having measures of working memory that mirror the way in which working memory operates in real experiences. For example, in this research, simple span tasks, like the digit span and Corsi block task, consistently did not predict posttest outcomes, but complex span tasks, like the operation span task, did. As prior literature has suggested, complex span tasks are much more similar to how our brains actually interact with incoming information in everyday situations (Engle et al., 1999; Waris et al., 2017). This literature demonstrates the need to measure individual differences in working memory in order to better understand learning, so it is important that there are many tasks that measure both working memory capacity and executive function that mirror real-life use of these factors.

Limitations and Future Directions

There are some notable limitations to this research that must be discussed. The first limitation is the lack of consistency in the measure of executive function that correlated with posttest scores in each of the experiments. Although each experiment showed that at least one of the measures of executive function used in the research had a significant relationship with posttest scores in lessons with distractors, it was not the same task each time (i.e., posttest score correlated with n-back task in Experiment 1, flanker task in Experiment 2, and Stroop and n-back tasks in Experiment 3). One potential reason for this was that the measures had low power due to how few trials each task had, so there was likely some variability in how accurate each participant's score was. As such, future research should investigate how better and more consistent measures of executive function predict posttest performance. Additionally, future research should try to identify if specific measures of executive function predict learning outcomes best.

Another limitation to this work is related to the length of the lesson. Across all three experiments, participants were given learning material that was presented to them in under 10 minutes. This makes it difficult for these results to generalize directly to real-world learning environments, like classrooms, as real-world learning experiences generally are much longer than 10 minutes. With longer durations of instruction and instructional material, the impacts of managing and holding incoming information may differ from what was found in these studies. It is important to understand how individual differences in managing and holding incoming information impact more real-world learning, as longer durations of learning time may make the impacts of working memory differences on learning more important or negate their effects.

An additional limitation of this work comes from the make-up of the samples used across the three studies. Almost all participants were between 18 and 25 years old and undergraduate students at University of California, Santa Barbara. As such, the findings from this research may not generalize to other groups of individuals. Particularly, the findings may not be the same for students who are younger, like those in K-12, where the integration of IVR into classes is becoming quickly widespread. More research needs to be done on this younger population in order to determine if the impacts of managing and holding incoming information on learning is similar for this group.

In addition to the research that should be conducted to help address the limitations of the series of experiments, there are also several other directions that future research should investigate. First, future research should investigate which specific features of an IVR lesson contribute to the learner's overall distraction in order to be able to more effectively use IVR technology in learning environments. As demonstrated in this research, being familiar with how to use VR does not seem to play a huge role in mitigating the distractions that are present in IVR lessons. Thus, to better attempt to mitigate the distractions created by IVR, it is necessary to identify what specifically is causing IVR to be distracting to learners.

Relatedly, more research needs to be conducted on how to measure distractions within an IVR lesson. As Experiment 2 demonstrated, the standard measures to understand the level of distraction learners experience during a lesson, like a question directly assessing how distracting a lesson was or a measure of extraneous processing while learning, may not be well equipped to measuring distraction that is occurring in an IVR lesson. When being asked about distractions, it may be the case that learners are not thinking about the many different ways their attention was not focused on the specific parts of the lesson while learning (e.g., distraction can occur if they: are focused on the wrong information within the lesson, are focused on using the VR device, are mind-wandering, etc.). As such, these current measures of distraction could be misleading. More research needs to be done to determine the best ways to accurately determine distraction in an IVR experience in order to have better findings about the distractibility of different IVR lessons.

Additionally, future research could enhance the use of IVR in learning by determining which types of experiences may best support learning within an IVR lesson. For example, some research has investigated how generative learning strategies, like summarizing, can be incorporated into an IVR lesson and benefit learning compared to a base IVR lesson (Parong & Mayer, 2018). However, there are only a couple of studies that presently have investigated this research and some of the findings show benefits of these strategies (i.e., Klingenberg et al., 2022; Parong & Mayer, 2018) while other studies do not show benefits (Elme et al., 2022; Parong & Mayer, 2021b). Much more research needs to be done to understand what other activities can make IVR more beneficial, especially with individual differences in mind.

Lastly, future research should work to create better measures of executive function and working memory capacity. Although there are many measures of executive function and working memory capacity, many of these tasks require many trials over a long period of time and work best in a controlled laboratory environment. Shorter versions of these tasks seem to have less reliable and less powerful outputs, as demonstrated with this set of experiments. For research on learning, it would be beneficial to have measures that are quick and easy to use in learning environments.

Conclusion

This set of three experiments aimed to understand how individual difference factors, specifically differences in ability to manage incoming information and ability to hold incoming information, impacted how well individuals could learn from a lesson with distractions. Across both multimedia video lessons and lessons in immersive virtual reality, a learner's ability to manage incoming information consistently predicted posttest scores when a lesson had distracting features; learners who were better at managing incoming information had better posttest scores compared to learners who were worse at managing incoming information. A learner's ability to hold new incoming information did not consistently predict posttest scores in any type of lesson. This research demonstrates the importance of considering individual difference factors in designing and implementing technology-based instruction.

References

Ackerman, P. L., & Lohman, D. F. (2006). Individual differences in cognitive functions. In P.A. Alexander & P.H. Winne (Eds.), *Handbook of educational psychology* (pp. 139-161). Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9780203053874-15

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

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2; pp. 89-195). Academic Press. https://doi.org/10.1016/s0079-7421(08)60422-3

Ayres, P., & Paas, F. (2007). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21*, 695-700. https://doi.org/10.1002/acp.1343

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-723. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8; pp. 47-89). Academic Press. https://doi.org/10.1016/s0079-7421(08)60452-1

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485-493. https://doi.org/10.1037/0894-4105.8.4.485

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science, 18*(2), 89-94. https://doi.org/10.1111/j.1467-8721.2009.01615.x

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*(1), 65-94. https://doi.org/10.1037/0033-2909.121.1.65

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*(4), 553-573. https://doi.org/10.1037/0033-2909.130.4.553

Begolli, K. N., Richland, L. E., Jaeggi, S. M., Lyons, E. M., Klostermann, E. C., & Matlen, B. J. (2018). Executive function in learning mathematics by comparison: Incorporating everyday classrooms into the science of learning. *Thinking & Reasoning, 24*(2), 280-313. https://doi.org/10.1080/13546783.2018.142306

Brown, T. E. (2009). ADD/ADHD and impaired executive function in clinical practice. *Current Attention Disorders Reports, 1*, 37-41. https://doi.org/10.1007/s12618-009-0006-3

Calvert, J., & Abadia, R. (2020). Impact of immersing university and high school students in educational linear narratives using virtual reality technology. *Computers & Education, 159*, 104005. https://doi.org/j.compedu.2020.104005

Carroll, J. B., & Maxwell, S. E. (1979). Individual differences in cognitive abilities. *Annual Review of Psychology, 30*, 603-640. https://doi.org/10.1146/annurev.ps.30.020179.003131

Chen, S.-K., Yeh, Y.-C., Hwang, F.-M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort-multioccasion study. *Learning and Individual Differences, 23*, 172-178. https://doi.org/10.1016/j.lindif.2012.07.021

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

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Theoretical and Review Articles, 12*, 769-786. https://doi.org/10.3758/BF03196772

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. [Doctoral dissertation, McGill University]. eScholarship@McGill.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismajatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42-100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466. https://doi.org/10.1016/s0022-537(80)90312-6

Diamond, A. (2012). Executive functions. *Annual Review of Psychology, 64*(1), 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Ekstrand, C., Jamal, A., Nguyen, R., Kudryk, A., Mann, J., & Mendez, I. (2018). Immersive and interactive virtual reality to improve learning and retention of neuroanatomy in medical students: A randomized controlled study. *CMAJ Open, 23*(6), E103-E109. https://doi.org/10.9778/cmajo.20170110

Elme, L., Jørgensen, M. L. M., Dandanell, G., Mottelson, A., & Makransky, G. (2022). Immersive virtual reality in STEM: Is IVR an effective learning medium and does adding self-explanation after a lesson improve learning outcomes? *Educational Technology Research and Development, 70*, 1601-1626. https://doi.org/10.1007/s11423-022-10139-3

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23, https://doi.org/10.1111/1467-8721.00160

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 28*(3), 309-331. https://doi.org/10.1037/0096-3445.128.3.309

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143-149. https://doi.org/10.3758/bf03203267

Fauville, G., Queiroz, A. C. M., Hambrick, L., Brown, B. A., & Bailenson, J. N. (2021). Participatory research on using virtual reality to teach ocean acidifcation: A study in the marine education community. *Enviornmental Education Research, 27*(2), 254-278. https://doi.org/10.1080/13504622.2020.1803797

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

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

Fiorella, L. & Mayer, R. E. (2022). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 185-198). Cambridge University Press. https://doi.org/10.1017/9781108894333

Flavia di Natale, A., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in k-12 and higher education: A 10-year systematic review of empirical research.

British Journal of Educational Technology, 51(6), 2006-2033. https://doi.org/10.1111/bjet.13030

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), 172-179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*(2), 201-225. https://doi.org/10.1037/0096-3445.137.2.201

Gilmore, C. K., & Papadatou- Pastou, M. (2009). Patterns of individual differences in conceptual understanding and arithmetical skill: A meta-analysis. *Mathematical Thinking and Learning, 11*(1-2), 25-40. https://doi.org/10.1080/10986060902583923

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

Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education, 8*, 1-32. https://doi.org/10/1007/s40692-020-00169-2

Han, I. (2020). Immersive virtual field trips in education: A mixed-methods study on elementary students' presence and perceived learning. *British Journal of Educational Technology, 51*(2), 420-435. https://doi.org/10.1111/bjet.12842

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*(1), 92-102. https://doi.org/10.1037/0022-0663.89.1.92

Hayes, A. F., & Coutts, J. J. (2020). Use Omega rather than Cronbach's Alpha for estimating reliability. But…, *Communication Methods and Measures, 14*(1). https://doi.org/10.1080/19312458.2020.1718629

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

Ilkowska, M., & Engle, R. W. (2010). Trait and state differences in working memory capacity. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook on individual differences in cognition: Attention, memory, and executive control* (p. 295-320). Spring Science + Business Media. https://doi.org/10.1007/978-1-4419-1210-7_18

Kalyuga, S., & Liu, T. C. (2015). Managing cognitive load in technology-based learning environments. *Educational Technology & Society, 18*(4), 1-8. https://www.jstor.org/stable/jeductechsoci.18.4.1

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169-183. https://doi.org/10.1037/0096-3445.130.2.169

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review, 27*, 317-326. https://doi.org/10.1007/s10648-015-9309-3

Kim, H. K., Park, J., Choi, Y., & Choe, M., (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics, 69*, 66-73. https://doi.org/10.1016/j.apergo.2017.016

Klingenberg, S., Fischer, R., Zettler, I., & Makransky, G. (2022). Facilitating learning in immersive virtual reality: Segmentation, summarizing, both or none? *Journal of Computer Assisted Learning, 39,* 218-230. https://doi.org/10.1111/jcal.12741

Kozhevnikov, M., Gurlitt, J., & Kozhevnikov, M. (2013). Learning relative motion concepts in immersive and non-immersive virtual environments. *Journal of Science Educational and Technology, 22*(6), 952-962. https://doi.org/10.1007/s10956-103-9441-0

Kwong See, S. T., & Ryan, E. B. (1995). Cognitive mediation of adult age differences in language performance. *Psychology and Aging, 10*(3), 458-468. https://doi.org/10.1037/0882-7974.10.3.458

Lee, S. H., Sergueeva, K., Catangui, M., & Kandaurova, M. (2017). Assessing Google Cardboard virtual reality as a content delivery system in business classrooms. *The Journal of Education for Business, 92*(4), 153-160. https://doi.org/10.1080/08832323.2017.1308308

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279-281. http://doi.org/10.1038/36846

MacLeod, C. M. (1991). Half a century of research on the Stroop Effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203. https://doi.org/10.1037/0033-2909.109.2.163

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop Effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Science, 4*(10), 383-391. https://doi.org/10.1016/s1364-6613(00)01530-8

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

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*, 937-958. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019b). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225-236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Mayer, R. E. (2021). *Multimedia Learning* (3rd ed). Cambridge University Press. https://doi.org/10.1017/9781316638088

Mayer, R. E. (2022). Cognitive theory of multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3nd ed., pp. 57-72). Cambridge University Press. https://doi.org/10.1017/9781108894333.008

Mayer, R. E., & Fiorella, L. (2022). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3nd ed., pp. 243-260). Cambridge University Press. https://doi.org/10.1017/9781108894333

Mayer, R. E., Makransky, G., & Parong, J. (2022). The promise and pitfalls of learning in immersive virtual reality. *International Journal of Human-Computer Interactions*, 1-10. https://doi.org/10.1080/10447318.2022.2108563

Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied, 8*, 147-154. https://doi.org/10.1037/10.1037/1076-898x.8.3.147

McDaniel, M. A., Anderson, J. L., Derbish, M. H., Morrisette, N. (2007). Testing the testing effect in classroom, *European Journal of Cognitive Psychology, 19*(4-5), 494-513. https://doi.org/10.1080/09541440701326154

Meltzer, L. & Krishnan, K. (2007). Executive function difficulties and learning disabilities: Understandings and misunderstandings. In L. Meltzer (Ed.), *Executive function in education* (pp. 77-105). The Guilford Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. https://doi.org/10.1037/h0043158

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167-202. https://doi.org/10.1146/annurev.neuro.24.1.167

Miller, K. M., Price, C. C., Okun, M. S., Montijo, H., & Bowers, D. (2009). Is the n-back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology, 24*(7), 711-717. https://doi.org/10.1093/arclin/acp063

Minnaert, A. E. (1999). Individual differences in text comprehension as a function of test anxiety and prior knowledge. *Psychological Reports, 84*(1), 167-177. https://doi.org/10.2466/PR0.84.1.167-177

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117-125. https://doi.org/10.1037/0022-0663.92.117

Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., & Shibasaki, H. (2004). The neural basis of executive function in working memory: An fMRI study based on individual differences. *Neuroimage, 21*(2), 623-631. https://doi.org/10.1016/j.neuroimage.2003.09.069

Paas, F., & Sweller, J. (2022). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 73-81). Cambridge University Press. https://doi.org/10.1017/9781108894333.009

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

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

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

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

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and instruction, 12*, 61-86. https://doi.org/10.1016/s0959-4752(01)00016-0

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164-171. https://doi.org10.1027/1015/5759/a000123

Rey, G. D. (2012). A review and a meta-analysis of the seductive detail effect. *Educational Psychology Review: Applied, 17*(2), 159-173. https://doi.org/10.1016/j.edurev.2012.05.003

Rhodes, S. M., Booth, J. N., Campbell, L. E., Blythe, R. A., Wheate, N. J., & Delibegovic, M. (2014). Evidence for a role of executive functions in learning biology. *Infant and Child Development, 23*(1), 67-83. https://doi.org/10.1002/icd.1823

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*. http://dx.doi.org/10.103/a0037559

Stepan, K., Zeiger, J., Hanchuk, S., Del Signore, A., Shrivastava, R., Govindaraj, S., & Iloreta A. (2017). Immersive virtual reality as a teaching tool for neuroanatomy. *International Forum of Allergy & Rhinology, 7*(10), 1006-1013. https://doi.org/10.1002/air.21986

Stoet, G. (2010). PsyToolkit – A software package for programming psychological experiments using Linux. *Behavioral Research Methods, 42*(4), 1096-1104. https://doi.org/10.3758/brm.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24-31. https://doi.org/10.1177/0098628316677643

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662. https://doi.org/10.1037/h0054651

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

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational technology research and development, 68*, 1-16. https://doi.org/10/1007/s11423-019-09701-3

Waris, O., Soveri, A., Ahti, M., Hoffing, R. C., Ventus, D., Jeaggi, S. M., Seitz, A. R., & Laine, M. (2017). A latent factor analysis of working memory measures using large-scale data. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01062

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

Wright, L., Lipszyc, J., Dupuis, A., Thayapararajah, S. W., & Schachar, R. (2014). Response inhibition and psychopathology: A meta-analysis of go/no-go task performance. *Journal of Abnormal Psychology, 123*(2), 429-439. https://doi.org/10.1037/a0036295

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology, 51*(6), 1991-2005. https://doi.org/10.1111/bjet.13023

Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). *Executive function: Implications for education.* National Center for Educational Research, Institute of Educational Science.

Appendices

Appendix A

Prequestionnaires

Experiment 1

Major:

GPA:

Age:

Gender:

Class level:
- o Freshman
- o Sophomore
- o Junior
- o Senior
- o Other (please specify)

Please rate your knowledge of fusion energy.
- o Very Low
- o Somewhat Low
- o Average
- o Somewhat High
- o Very High

Please select each of the things that apply to you:
- o I have taken a class that has discussed fusion energy.
- o I understand how the sun produces energy.
- o I am concerned with how alternative energy systems work.
- o I know the difference between fusion and fission.
- o I have taken a class that has discussed nuclear energy.
- o I am interested in alternative types of energy.
- o I am interested in how the sun creates energy.

Experiments 2 & 3

Major:

GPA:

Age:

Gender:

Class level:

- Freshman
- Sophomore
- Junior
- Senior
- Other (please specify)

Please rate your knowledge of ocean acidification.

- Very Low
- Somewhat Low
- Average
- Somewhat High
- Very High

Please select each of the things that apply to you:

- I have taken a class that has discussed greenhouse gases.
- I have seen the effects of ocean acidification first-hand.
- I am a member of a conservation group (such as Sierra Club).
- I like to watch the National Geographic cable channel.
- I consider myself to be an environmentalist.
- I know how ocean acidification works.
- I have taken conscious efforts to reduce my carbon footprint.
- I have learned about how climate change occurs.
- I am concerned about ocean reefs.

How prone are you to motion sickness?

- Not At All
- Very Little
- Moderate
- Very Prone

Appendix B

Grading Rubrics

Experiment 1

Question 1: How does fusion work in the sun?

- o Sun cheats with high pressure instead of high heat
- o Strips electrons from nuclei
- o Overcomes repelling nuclei
- o Collisions occur
- o Nuclei get merged together
- o Energy is released

Question 2: Why is fusion energy so efficient?

- o Hydrogen is abundant/easy to find (in sea water)
- o Safe reaction
- o Creates a lot of energy
- o Very little waste is created/cleaner energy source

Question 3: Why can the sun create fusion energy without having a high enough temperature?

- o High pressure instead of heat
- o Pressure can squeeze together the nuclei

Question 4: What are the pros of using fusion energy on Earth?

- o Unlimited energy from fusion/efficient
- o Very limited affect on environment/ clean energy
- o Easy to acquire most materials
- o Safer than other forms of energy

Question 5: What is one possible environmental reason to not use fusion energy?

- o Tritium is radioactive (if used as a fuel source)
- o Tritium can leak into the environment as a byproduct

Question 6: What are the cons of using fusion energy?

- o Potential environmental concerns
- o Need to harvest some materials, that may be harder to get
- o Expensive to build now
- o Not a proven technology
- o Can use money on other types of proven alternative energy

Question 7: What happens if the confinement of a fusion reactor fails?

- o Plasma expands
- o Plasma cools off
- o Plasma needs to be moving fast or very hot to make reaction work
- o No reaction would occur

Question 8: Why does the reaction stop if the confinement of a fusion reactor fails?

- o Hot temperature/high pressures required for fusion
- o Plasma cools
- o Plasma loses pressure (it expands)
- o Nuclei repel one another again

Experiments 2 and 3

Question 1: Explain how ocean acidification occurs.
- CO2 is in the atmosphere
- CO2 and H2O interact
- Creation of acid/acidification
- Specifically, carbonic acid is created
- Carbonic acid increases the acidity

Question 2: If car's exhaust released a different greenhouse gas, like methane (CH4), could ocean acidification still occur due to car fumes?
- No
- CO2 + H2O = carbonic acid only
- H2O and CH4 do not combine to create carbonic acid specifically

Question 3: What steps would need to be taken to make the ocean less acidic?
- Reduce greenhouse gas producing items
- Specifically, reduce CO2
- This lessens the interaction with water
- Creates less carbonic acid/less acid/less ocean acidity

Question 4: What molecules combine to increase ocean acidification? What molecule do they create?
- CO2 (carbon dioxide)
- H2O (water)
- H2CO3 (carbonic acid)

Question 5: How does an abundance of CO2 in the atmosphere affect sea life?
- Interaction of H2O and CO2
- More carbonic acid in the ocean
- Acidic ocean/ ocean acidification occurs
- Acid is harmful to sea life
- Sea life moves or dies, unlivable conditions for sea life

Question 6: Why were their sea snails in the first area of the reef but not the second area?
- First area = less CO2 absorbed
- First area = less/not acidic
- First area = not corrosive
- Second area = more CO2 absorbed
- Second area = acidic/ acid present
- Second area = corrosive area kills off the sea life

Question 7: How might ocean acidification affect humans?
- Kills sea life/less sea life
- Food chain can collapse/ecosystem collapse
- Less fish in the sea means less fish for humans to consume

Question 8: How can marine scientists figure out if an area of a reef is healthy?
- Species count
- More sea life = healthy reef
- No/little sea life = unhealthy reef

Appendix C

Postquestionnaire

Experiment 1

Experience Questions
1. I enjoyed this lesson.
2. The topic of this lesson was interesting to me.
3. I would like to learn from more lessons like this.
4. I felt as though they way this lesson was taught was effective for me.
5. I felt motivated to understand the material.
6. The video lesson was distracting. (only in Studies 1b and 1c)

Cognitive Load Questions[*]
7. It was hard to pay attention during the lesson. (extraneous processing)
8. I tried to remember the information in the order presented. (essential processing)
9. I felt distracted during the lesson. (extraneous processing)
10. I was trying to make sense of the material. (generative processing)
11. I was working to memorize the information. (essential processing)
12. I was trying to make connections between the material and things I already know. (generative processing)
13. I was working on understanding the lesson. (generative processing)
14. My mind was not on the lesson. (extraneous processing)
15. I was trying to learn the main facts from the lesson. (essential processing)

Other Questions
16. Are you color blind?[^]
17. Please provide any additional comments about this lesson below. (Open-ended question)
18. Did anything not work for you during this lesson?[^]

Experiment 2

Experience Questions[*]
1. I enjoyed this lesson.
2. The topic of this lesson was interesting to me.
3. I would like to learn from more lessons like this.
4. I felt as though they way this lesson was taught was effective for me.
5. I felt motivated to understand the material.
6. The video lesson was distracting.

Cognitive Load Questions[*]
7. It was hard to pay attention during the lesson. (extraneous processing)

8. I tried to remember the information in the order presented. (essential processing)
9. I felt distracted during the lesson. (extraneous processing)
10. I was trying to make sense of the material. (generative processing)
11. I was working to memorize the information. (essential processing)
12. I was trying to make connections between the material and things I already know. (generative processing)
13. I was working on understanding the lesson. (generative processing)
14. My mind was not on the lesson. (extraneous processing)
15. I was trying to learn the main facts from the lesson. (essential processing)

Prescence Questions[*]
1. I felt that I was able to control events in the environment.
2. The environment was responsive to actions that I performed.
3. My interaction with the environment seemed very natural.
4. The visual aspects of the environment were really present and moving around me.
5. I felt like the objects in the environment were really present and moving around me.
6. I felt like I was really in the environment.
7. I felt like I was immersed (or included in) and interacting with the environment.

Cyber-sickness Questions[~]
1. General Discomfort (oculomotor question)
2. Fatigue~ (oculomotor question)
3. Eye Strain~ (oculomotor question)
4. Difficulty Focusing~ (oculomotor question)
5. Headache~ (disorientation question)
6. Fullness of Head~ (disorienting question)
7. Blurred Vision~ (disorienting question)
8. Dizzy (Eyes Closed)~ (disorienting question)
9. Vertigo~ (disorienting question)

Other Questions
1. Are you color blind?[^]
2. Please provide any additional comments about this lesson below. (Open-ended question)
3. Did anything not work for you during this lesson?[^]
4. How much experience did you have with VR prior to participating in this study?[#]

Experiment 3

Experience Questions[*]
1. I enjoyed this lesson.
2. The topic of this lesson was interesting to me.
3. I would like to learn from more lessons like this.
4. I felt as though they way this lesson was taught was effective for me.

5. I felt motivated to understand the material.
6. The video lesson was distracting.
7. The instructions for the lesson I saw distracted me from the lesson.
8. Interacting with the environment in the lesson distracted me from the content.
9. Being immersed in the environment in the lesson distracted me from the content.

Cognitive Load Questions[*]
1. It was hard to pay attention during the lesson. (extraneous processing)
2. I tried to remember the information in the order presented. (essential processing)
3. I felt distracted during the lesson. (extraneous processing)
4. I was trying to make sense of the material. (generative processing)
5. I was working to memorize the information. (essential processing)
6. I was trying to make connections between the material and things I already know. (generative processing)
7. I was working on understanding the lesson. (generative processing)
8. My mind was not on the lesson. (extraneous processing)
9. I was trying to learn the main facts from the lesson. (essential processing)

Presence Questions[*]
1. I felt that I was able to control events in the environment.
2. The environment was responsive to actions that I performed.
3. My interaction with the environment seemed very natural.
4. The visual aspects of the environment were really present and moving around me.
5. I felt like the objects in the environment were really present and moving around me.
6. I felt like I was really in the environment.
7. I felt like I was immersed (or included in) and interacting with the environment.

Cyber-sickness Questions[~]
1. General Discomfort (oculomotor question)
2. Fatigue~ (oculomotor question)
3. Eye Strain~ (oculomotor question)
4. Difficulty Focusing~ (oculomotor question)
5. Headache~ (disorientation question)
6. Fullness of Head~ (disorienting question)
7. Blurred Vision~ (disorienting question)
8. Dizzy (Eyes Closed)~ (disorienting question)
9. Vertigo~ (disorienting question)

Job Simulator Experience Questions[*]
1. I enjoyed playing Job Simulator.
2. I felt that playing Job Simulator was a waste of time.
3. I learned how to use the VR better because of Job Simulator.
4. Playing Job Simulator distracted me from learning the material presented in the lesson.

Other Questions

19. Are you color blind?[^]
20. Please provide any additional comments about this lesson below. (Open-ended question)
21. Did anything not work for you during this lesson?[^]
22. How much experience did you have with VR prior to participating in this study?[#]

[*]Scale = 1 – Strongly Disagree, 2 – Somewhat Disagree, 3 – Neutral, 4 – Somewhat Agree, 5 – Strongly Agree

[^]Scale = 1 – Yes, 2 – No

[~]Scale = 1 – Not At All, 2 – Slightly, 3 – Moderately, 4 – Very

[#]Sclae = 1 – No experience with VR, 2 – Very little experience with VR, 3 – Moderate amount of experience with VR, 4 – A lot of experience with VR

Appendix D

Experiments 2 & 3 IVR Lesson Scripts

IVR Version:

Climate change has begun. It is already hurting our planet. Many of us have seen the toll on cities and coast lines already. But very few people witness the damage below the ocean's surface. Now what if you had a crystal ball, and that crystal ball showed you exactly what the world's oceans would look like in a world affected by climate change. Walk up and touch the crystal ball to experience it for yourself.

Look at the grey car in front of you. Bend down and touch the exhaust pipe. These are CO2 molecules. Take a step back and watch them spew from the exhaust pipe. Humans release over 22 million tons of CO2 into the atmosphere each day. Pay special attention to the larger molecule now in front of you. Let's see where it goes.

Look out onto the water. The molecule that you see floating on the ocean's surface is a water molecule known as H2O. Now, remember that CO2 molecule from the car? Look up and watch as it floats down from the sky. To watch it fall into the water, push the CO2 molecule with your hand. Observe the chemical reaction that occurs when CO2 combines with H2O to create carbonic acid, or H2CO3. This process is call ocean acidification. If enough carbonic acid is created, the sea water becomes corrosive. Now look around as many CO2 molecules from the air get absorbed into the ocean. Since the industrial revolution, the ocean has absorbed roughly ¼ of the CO2 produced by burning fossil fuels. You will now travel to a special site where scientists have made a breakthrough discovery.

We are now underwater on a rocky reef of the coast of Naples, Italy. What makes this reef special is the natural, underwater vents that spew CO2 onto the rocky reefs, making this water more acidic. The reef serves as a crystal ball through which scientists can see the effects of increased acidity on ocean ecosystems.

People learn best from firsthand experiences. We've used computer graphics to allow you to interact with the reef. Look around. Notice the vibrant colors and different animal species around you. As a marine scientist, it is your job to measure the health of this reef by doing a species count. Today you will focus on sea snails. Look down at your right hand and rotate it until you can see the snail. This is the type of healthy sea snail you'll be looking for today. Look around for the flags in the blue basket. Reached down and touch the flags with your hand to pick one up. Now that you have a flag, find a snail. Reach out with the flag in hand and touch the snail to put the flag beside it. Continue to pick up flags from the pile and place them on the snails surrounding you. Try to be quick. Time is up. We will now move to a more acidic part of the reef.

Observe the changes that have occurred as acidity has increased. Reach out and touch one of the streams of bubbles rising from the sea floor. Remember that the vents here are releasing CO_2 which is making the water more acidic. You study this part of the reef to predict the future. How human emissions will affect our world's oceans and the species that live within them. Look at your right palm. Notice how acidity has eroded the sea snail's shell. Now take a moment to walk around and look for sea snails in this area. Couldn't find any? That's because there are no living sea snails here. They can't survive in this environment. Ocean acidification will severely impact all shelled species, including oysters, clams, corals, and certain kinds of plankton. Without these species, the entire food web can collapse. All of our oceans will look like this Mediterranean reef unless we reduce our CO_2 emissions. But it's not too late.

There are actions you can take to combat ocean acidification. Keep learning about climate change and share that knowledge with others. Take steps to reduce your own carbon footprint. You can also support ocean acidification research and urge decision makers to provide funding to the National Oceanic and Atmospheric Administration, which helps communities and businesses understand the risk posed by acidification. The future of our Earth is in your hands.

Slideshow Lesson

Climate change has begun. It is already hurting our planet. Many of us have seen the toll on cities and coast lines already, but very few people witness the damage below the ocean's surface. Now what if you had a crystal ball? And that crystal ball showed you exactly what the world's oceans would look like in a future affected by climate change?

Look at the grey car in front of you on the screen. From the exhaust pipe carbon CO_2 molecules are being spewed. They then go up into the air. Humans release over 22 million tons of CO_2 into the atmosphere each day. Let's see where these molecules go.

Look out onto the water in image 1. The molecule floating on the ocean's surface is a water molecule known at H_2O. Now, remember that CO_2 molecule from the car? Look at the CO_2 molecule come down to the ocean in image 2. There is a chemical reaction that occurs when CO_2 combines with H_2O to create carbonic acid, or H_2CO_3, shown in image 3. This process is called ocean acidification. If enough carbonic acid is created, the sea water becomes corrosive.

As you can see, there are many CO_2 molecules in the air that get absorbed into the ocean. Since the industrial revolution, the ocean has absorbed roughly 1/4 the of the CO_2 produced by burning fossil fuels. We'll now travel to a special site where scientists have made a breakthrough discovery.

We are now underwater on a rocky reef of the coast of Naples Italy. What makes this reef special is the natural underwater vents that spew CO2 onto the rocky reefs, making this water more acidic. The reef serves as a crystal ball through which scientists can see the effects of increased acidity on ocean ecosystems.

Look around at the images. Notice the vibrant colors and different animal species you see. As a marine scientist, your job would be to measure the health of this reef by doing a species count.

Today you will focus on sea snails. Look at this image. This is a healthy sea snail marine scientists look for.

Look at these images. You will see that everywhere there is a flag indicates where a sea snail is found.

Now you have moved to a more acidic part of the reef. Observe the changes that have occurred as acidity has increased in these images. The bubble vents in the images are releasing CO2 what is making the water more acidic. Marine scientists study this part of the reef to predict the future; how human emissions will effect our world's oceans and the species that live within them.

Look at the shell now. Notice how acidity has corroded the sea snails shell.

Now take a moment to look for flags indicating sea snails in this area. Couldn't find any? That's because there are no living sea snails here. They can't survive in this environment.

Ocean acidification will severely impact all shelled species including oysters, clams, corrals, and certain types of plankton. Without these species, the entire food web can collapse.

All of our oceans will look like this Mediterranean reef unless we reduce our CO2 emissions. But it's not too late. There are actions you can take to combat ocean acidification. Keep learning about climate change and share that knowledge with others. Take steps to reduce your own carbon foot print. You can also support ocean acidification research and urge decision makers to provide funding to the National Oceanic and Atmospheric Administration which helps communities and businesses understand the risks posed by acidification. The future of our earth is in your hands.

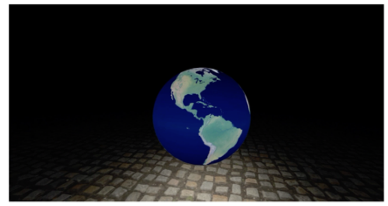
Appendix E

Slides from Slideshow Lesson in Experiment 2

Slide 1

The Stanford Ocean Acidification Experience

Slide 2

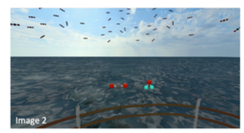


- Climate change = hurting the planet
- This lesson will show you how the world's oceans would look in a future affected by climate change

Slide 3

- CO_2 = released from a exhaust pipe and goes into the air
- Humans release 22 million tons of CO_2 into the atmosphere each day

Slide 4



- The molecule floating on the ocean is a water molecule, H_2O (image 1)
- CO_2 + H_2O (image 2) = carbonic acid, H_2CO_3 (image 3)
- called ocean acidification
- Carbonic acid = corrosive sea water

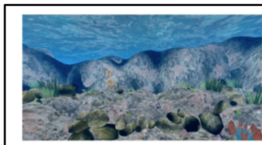
Slide 5

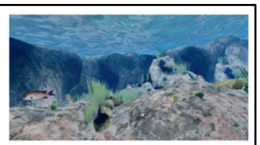
- CO_2 molecules from the air get absorbed into the ocean

Slide 6

- rocky reef off the coast of Naples, Italy
- Natural underwater vents spew CO_2 - water becomes more acidic
- Used to see the effects of increased acidity on ocean ecosystems

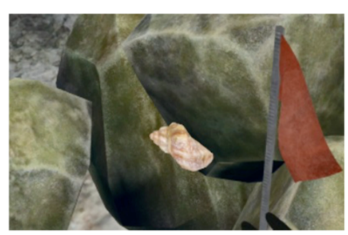
Slide 7





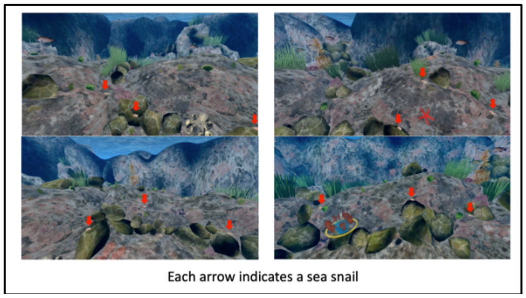
- Marine scientists measure the health of the reef by doing a species count, like counting sea snails

Slide 8

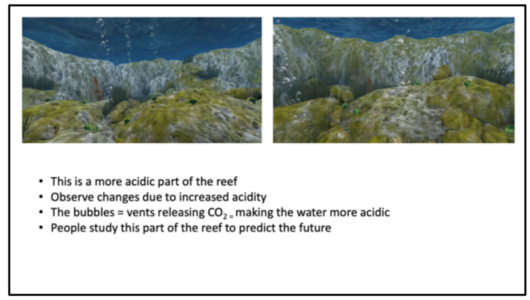


Healthy sea snail

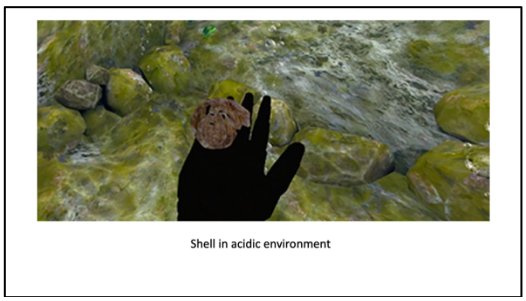
Slide 9



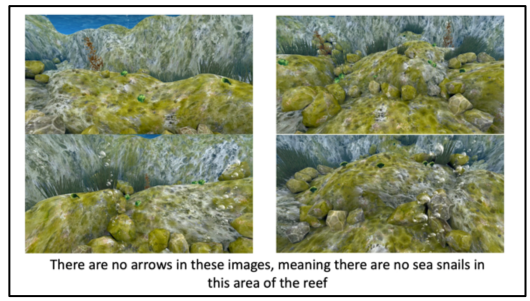
Slide 10



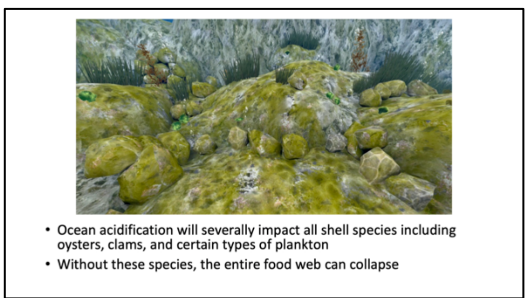
Slide 11



Slide 12



Slide 13



Slide 14