2023

High School Science Students' Cognitive Load Using Virtual Reality Compared to Traditional Instruction

Maria Carmen R. Lagalante
Walden University

Walden University

College of Education and Human Sciences

This is to certify that the doctoral dissertation by

Maria Carmen R. Lagalante

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Darci J. Harland, Committee Chairperson, Education Faculty
Dr. Charlotte Redden, Committee Member, Education Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2023

Abstract

High School Science Students' Cognitive Load Using Virtual Reality Compared to

Traditional Instruction

by

Maria Carmen R. Lagalante

MA, Azusa University, 2012

BA, Thomas Aquinas College, 1999

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

August 2023

Abstract

Cognitive load is the effort needed to process and store information in memory and can be measured via subjective, physiological, and performance methods. Virtual reality learning environments (VRLEs) enhance science aptitude and motivate students to pursue science careers. Cognitive load is divided into three types: intrinsic, extraneous, and germane. Using cognitive load theory as the framework, the problem addressed through this study was that it is not yet fully understood what the effect of VRLEs is on students' cognitive load, which can hinder their learning if it is too high. Secondary science education comprises many complex topics with significant levels of intrinsic cognitive load. Discovering if VRLEs reduce intrinsic cognitive load without increasing extraneous load, leaving more room for germane cognitive load to aid student processing is crucial. The purpose of this quantitative quasi-experimental study was to determine the difference in cognitive load as measured with the Mental Effort Survey (MES) between high school students who used a VRLE during science instruction for one lesson and students who did not use VRLEs. For this nonequivalent control group design, the data points were derived from high school students who completed Leppink's MES after a science lesson conducted during the 2021–2022 school year in a private high school in the southwestern United States. The means for each score of the two groups were compared using two-tailed *t* tests. Results showed a significant decrease in the intrinsic and extraneous load and a significant increase in germane load for the VRLE group. For positive social change results can inform stakeholders about the use of VRLEs and may contribute to academic success, secondary school graduation rates, and employment rates.

High School Science Students’ Cognitive Load Using Virtual Reality Compared to Traditional Instruction

by

Maria Carmen Lagalante

MA, Azusa Pacific University, 2012

BA, Thomas Aquinas College, 1999

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

August 2023

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Virtual reality learning environments (VRLE) are computer-created, immersive, interactive 3D worlds accessed via a head-mounted display (Parong & Mayer, 2018). VRLEs are increasingly being used in high school science courses to provide students with immersive, interactive experiences to enhance their understanding of complex scientific concepts (Bogusevschi et al., 2020; Khotimah et al., 2021; Lamb et al., 2018; Makransky & Lilleholt, 2018). These environments can simulate real-world scenarios and provide students with hands-on learning opportunities that can be difficult or impossible to achieve through traditional methods (Cheng & Tsai, 2019; Gielstra et al., 2021; Huang & Liaw, 2018; Kenna & Potter, 2018; Lee et al., 2021). Additionally, VRLEs can be used to present scientific information in a visually engaging way that can hold students' attention and foster their interest in the subject (Astuti et al., 2020; Boda & Brown, 2020; Hu Au & Lee, 2017; Huang & Liaw, 2018). The use of VRLEs in science education is still in its early stages, but early results from the published literature have suggested that it has the potential to be a valuable tool for engaging students and enhancing their learning outcomes (Bogusevschi et al., 2018; Hatchard et al., 2019; Lamb et al., 2018). However, more research is needed to study how VRLEs affects students' cognitive processes.

Cognitive load theory (CLT) provides a way to assess how VRLEs affects student learning. Cognitive load refers to the mental effort required to perform a task and can be quantified using reaction time, accuracy, and subjective ratings (Sweller et al., 2019). In high school science education, cognitive load can be influenced by various factors, such as the difficulty of the content, the teaching methods used, and the student's prior knowledge (Martin et al., 2020, 2021). High levels of cognitive load can make learning more difficult and reduce students' motivation (Sweller et al., 2019). Effective science education should minimize cognitive load and balance it with an appropriate level of challenge to promote deep learning and understanding.

VRLE technology can potentially increase access to educational opportunities for students who may not be able to participate in certain hands-on learning activities due to geographic location, physical abilities, or other constraints. For example, the use of VRLEs in high school science education offers the possibility of virtual field trips to locations that are otherwise unreachable by students (Cheng & Tsai, 2019; Lin et al., 2011; Petersen et al., 2020). This technology allows students to explore inaccessible locations, providing a unique and interactive learning experience. Another example is the expediency of virtual labs, which has resulted in the growth of online science courses and advanced lab sciences in rural areas, enabling students in these areas to have access to cutting-edge science education (Gielstra et al., 2021; Kenna & Potter, 2018; Seifan et al., 2020). Teachers and administrators can better understand the potential benefits and drawbacks by studying the cognitive load associated with using VRLEs in science instruction.

Background

It has been reported from various studies that VRLEs can be effective for learning; however, it has also been seen to depend on various factors, such as the student's learning style (Huang et al., 2020), spatial abilities (Lee & Wong, 2014), and gender (Ariali & Zinn, 2021; Ibili & Billinghurst, 2019). Although many studies have concluded that VRLEs create a sense of presence, researchers are divided on how it affects learning outcomes. Some studies have shown a correlation between an increased sense of presence and increased learning outcomes (Baceviciute et al., 2021; Lee et al., 2010; Sun et al., 2019), and others have shown the opposite (Makransky, Terkildsen, et al., 2019; Parong & Mayer, 2021). Further examination is needed to understand VRLEs' contribution to learning beyond the sense of presence.

The results of studies examining the effects of CLT in VRLEs have also been inconclusive. Some researchers have shown that adding static images to VRLEs can decrease cognitive load (Shin & Park, 2019), while others have shown no significant difference (Nelson et al., 2016). Another cognitive effect applied in VRLEs is the segmentation of information into smaller parts, which has been shown to reduce cognitive load in VRLEs just as in other media (Parong & Mayer, 2018).

To better understand these mixed results, it is crucial to measure the three types of cognitive load (i.e., intrinsic, extraneous, and germane) separately. The goal of an instructional tool is to lower intrinsic and extraneous load and increase the germane load while avoiding cognitive overload (Sweller, 2020). Several researchers attempted to measure these types of cognitive load separately, but they had differing results (Parong & Mayer, 2021; Vesga et al., 2021). VRLEs have the potential for effective learning, but more research that measures the three types of cognitive load is needed to fully understand their influence and how to optimize them for better learning outcomes.

VRLE use in high school science classrooms has significantly increased since 2018, with the increased ability for access in schools (Zhang et al., 2020). One of the benefits is an improved scientific attitude (Astuti et al., 2020; Garduño et al., 2021), which is a crucial aspect because as student interest in a subject increases, there is a tendency for higher learning achievements (Boda & Brown, 2020). The reasons for this increase include the ability to make intangible and inaccessible concepts tangible (Nersesian et al., 2019a; Petkov et al., 2019), the provision of a platform for embodied learning, and fostering an attitude that is open to inquiry and problem solving (Astuti et al., 2020). VRLE use in science classes has, in some cases, led to higher test scores and improved retention of knowledge with younger populations (Bogusevschi et al., 2018; Jitmahantakul & Chenrai, 2019; Lai et al., 2022; Liou & Chang, 2018; Nersesian et al., 2019b; Southgate, 2019). However, some studies with college-aged students showed that the VRLE group did not perform better than the traditional teaching group (Meyer et al., 2019; Parong & Mayer, 2018). There is a need for studies focusing on the cognitive effects VRLEs have on the high school population.

High school science education is an important area of concern for governments globally (Uçar & Sungur, 2017). Science education has been linked to economic progress and growth (Aguilera & Perales-Palacios, 2020; Murphy et al., 2018; Sadler et al., 2013). Yet, students worldwide are losing interest in science and scoring poorly in national science assessments (Achor et al., 2019; Aguilera & Perales-Palacios, 2020; Bal-Incebacak et al., 2019; Etobro & Fabinu, 2017; Hidayati et al., 2020; Lodge, 2021; Nidup et al., 2021; Santi & Gorghiu, 2019; Uçar & Sungur, 2017). This issue is likely related to the inherent complexity of science subjects, which pose a high level of intrinsic cognitive load that can easily lead to cognitive overload and prevent students from retaining information (Achor et al., 2019; Kokkonen et al., 2022; Steier & Kersting, 2019). Research in this area focuses on the inherent complexity of learning science (Achor et al., 2019; Aguilera & Perales-Palacios, 2020; Etobro & Fabinu, 2017; Hidayati et al., 2020; Santi & Gorghiu, 2019), reasoning schemas needed to learn science (Bal-Incebacak et al., 2019; Sadler et al., 2013; Santi & Gorghiu, 2019; Wei et al., 2021; Zhang, 2019), and interventions to decrease the cognitive load (Becker et al., 2020; Jitmahantakul & Chenrai, 2019; Lardi & Leopold, 2022; Saw, 2017; Weng et al., 2018). The current study was needed because it helps inform stakeholders as to whether one of the benefits of the innovation of VRLEs is that they reduce secondary students' cognitive load. A better understanding of the relationship between VRLEs and cognitive load may be used to help improve students' understanding of science.

Problem Statement

The problem addressed through this study was that it is not yet fully understood what the effect of VRLEs is on students' cognitive load, which can hinder their learning if it is too high. VRLEs are becoming a popular innovative learning mode to use in K–12 institutions. In the Horizon Report for K-12 Education, the New Media Consortium and Consortium for School Networking (2017) predicted that VRLEs could reach 15 million learners by 2025, and the Perkins Coie (2018) survey rated education the second most suitable use of virtual reality (VR). Multiple studies have shown that VRLEs improve students' motivation and engagement (Akman & Çakır, 2019; Cai et al., 2017; Cheng &

Tsai, 2019; Keller et al., 2018). In a literature review, Hu Au and Lee (2017) identified VRLEs as a good tool for teaching 21st-century skills, such as empathy, creativity, abstract thinking, and the ability to visualize complex concepts. VRLEs are used by science and technology instructors because they provide a way to mimic the natural world, allowing students to explore and interact with it (Huang & Liaw, 2018).

However, little is known about the effects of VRLEs on high school students learning science, specifically concerning cognitive load. Paas and Van Merrienboer (1994) warned that high cognitive load adversely affects learning, explaining that learning is impaired if the total processing time exceeds working memory capacity. Studies have shown that VRLEs improved eighth graders' science learning outcomes (Cai et al., 2017), high school students' higher order cognitive thinking (Southgate, 2019), and high school students' geoscience test scores (Jitmahantakul & Chenrai, 2019). However, none of these studies focused on cognitive load directly, even though the high cognitive load has been a concern related to this technology's use for educational purposes (Makransky, Terkildsen, et al., 2019). In another study, when high school students used VR, their learning style did not influence how well they learned, but it did impact a student's cognitive load (Huang et al., 2020). Huang et al. (2020) suggested additional studies on cognitive load and VRLEs are needed and specifically suggested the use of quasi-experimental designs. There is a gap in published studies that measure how VR environments affect science students' cognitive load.

The study is significant to the discipline of educational technology because VRLEs are a new technology being used in the classroom. This study extends the literature regarding how this technology affects the learning processes of students learning science.

Purpose of the Study

The purpose of this quantitative quasi-experimental study was to determine the difference in cognitive load as measured with the Mental Effort Survey (MES) between high school students who used a VRLE during science instruction for one lesson and students who did not use a VRLE.

Research Question and Hypotheses

To address the problem and purpose of this study, the following research questions and corresponding hypotheses guided the study:

RQ1: What is the difference in intrinsic load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE?

$H_0$1: There is no difference in intrinsic load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

$H_1$1: There is a statistical difference in intrinsic load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

RQ 2: What is the difference in extraneous load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE?

$H_0 2$: There is no difference in extraneous load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

$H_2 2$: There is a statistical difference in extraneous load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

RQ 3: What is the difference in germane load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE?

$H_0 3$: There is no difference in germane load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

$H_3 3$: There is a statistical difference in germane load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

Theoretical Framework

I used the CLT introduced in 1988 by John Sweller as the theoretical framework for this study. Sweller et al. (2019) suggested that CLT helps explain the human cognitive processes that consist of working memory and long-term memory and that processing issues occur because of the limited capacity of the working memory. Cognitive load consists of allocating mental resources required to move new information through the working memory and into the long-term memory. Sweller (1988) broke it into three types: intrinsic, extraneous, and germane. The intrinsic load comes from the complexity of the task and the amount of a student's prior knowledge (Leppink et al., 2013). The external load derives from sensory inputs that are not beneficial to learning (Leppink et al., 2013). The germane load arises from instructional material processing (Leppink et al., 2013). I provide more detail on the three types of cognitive load in Chapter 2.

Sweller (2020) distinguished primary general knowledge that is nonteachable and secondary subject-specific knowledge, which is teachable. In the CLT, Sweller provided guidelines for developing instructional designs to teach secondary knowledge, entering the working memory from visual and auditory channels. Students do not learn new information if the material does not reach long-term memory, where the data are stored for later use (Sweller, 2020). Mayer and Moreno (2003) applied the principles of CLT to the design of multimedia materials, finding that the use of extraneous material that requires processing by the visual or auditory channels can cause an external cognitive overload of the working memory. Instructional designers, especially of multimedia lessons, need to limit the use of the sensory channels to material necessary for instruction (Clark & Mayer, 2016). Multimedia designers need to ensure that learning activities lessen the intrinsic load, do not cause external cognitive overload, and focus all mental effort on the germane load so deep learning can occur (Sweller, 2020).

VR is the newest form of multimedia design, and many studies have examined how VR affects students' cognitive load (Chen, 2006; Dan & Reiner, 2017; Huang et al., 2020; Makransky, Terkildsen, et al., 2019). In the current study, I used CLT to examine the cognitive load variables of high school science students who learn content in a VRLE compared to traditional instruction. In this study, I used the validated quantitative survey, the self-reported Leppink MES, that examines the three cognitive load types identified in the CLT: intrinsic, extraneous, and germane. Therefore, the theory aligned with the instrument I used, which helped me interpret the results because they aligned well with the study's purpose and research questions.

Nature of the Study

One way to discern if VR environments cause cognitive overload is to measure the mental effort experienced by the students during the lesson (Leppink et al., 2013; Morrison et al., 2014; Paas & Van Merriënboer, 1994). Additionally, if cognitive overload exists, students cannot retain information learned and tend to score poorly on unit tests. Liou et al. (2017) used CLT to research the differences in learning outcomes between augmented reality (AR) and VR. Shin and Park (2019) employed CLT to show that 3D learning spaces reduced the cognitive load associated with 2D animations. In the current study, I used CLT to examine the cognitive load variables of high school science students who learned content in a VRLE compared to traditional instruction. By using the scores from the Leppink MES, I was able to examine the three types of cognitive load identified by CLT: intrinsic, extraneous, and germane.

I used a quantitative research paradigm for this study. A quantitative design can be justified for studying cognitive load because it allows for collecting and analyzing numerical data (see Burkholder et al., 2016). By using quantitative methods, researchers can obtain precise and objective data on cognitive load, which can be analyzed and compared across different conditions or group (DeLeeuw & Mayer, 2008; Paas et al., 1994; Paas & Van Merriënboer, 1994; Parong & Mayer, 2021). This analysis enables the researcher to identify patterns and relationships in the data and make more informed conclusions about the factors contributing to cognitive load. Furthermore, quantitative designs can also provide a level of standardization and control that is difficult to achieve with qualitative methods (Coleman, 2022), which can be particularly important in studying cognitive load because it is a complex and multidimensional construct (Sweller, 2011). Using a quantitative approach can provide valuable insight into the nature of cognitive load and its determinants, making it a fitting approach for studying this topic.

For the research design, I used archival data collected from MES data (see Leppink et al., 2013) from two groups of high school students: one who was taught in a VRLE in science class and the other who did not. Students self-assessed the cognitive load experienced while using the VRLE by taking a mental effort survey developed by Leppink et al. (2013). Students who learned the same science concept, not using a VRLE, also took the MES. The literature surrounding CLT has shown that the intensity of effort required to complete a task is also a reliable measurement of cognitive load (Paas et al., 2003). The MES has several questions regarding the factors that can make a task easier or harder to complete, such as the environment, amount of time spent, perceived difficulty, and ease of use of instruments, and the questions use a Likert-type scale ranging from 1 to 9 (Paas et al., 2003). I conducted this quasi-experimental study at one local high school in the southwestern United States to examine students' cognitive load who used a VRLE for a science lesson compared to students who learned the same science content with traditional instruction. The study site high school had 24 VR headsets donated, and the administration granted permission for their use at teachers' discretion. Some science teachers use these headsets for some lessons, but others do not, so that the comparison groups were natural, not assigned. I used the results from the MES, which has 10 questions total. Questions 1, 2, and 3 measure intrinsic cognitive load; Questions 4, 5, and 6 measure the extraneous load; and Questions 7, 8, 9,10 measure germane load. I used the survey results to answer the research questions by comparing the results of the two groups. Use of a two-sample *t* test was appropriate because the research questions required determining if the VRLE group's mean was significantly greater than or less than the traditionally taught group (see Frankfort-Nachmias & Leon-Guerrero, 2018).

Definitions

The following terms served to inform the study:

Cognitive load: The allocation of mental resources required to move new information through the working memory and into the long-term memory (Sweller, 1988). Measured by subjective rating scales, physiological measures, performance measures, or neuroimaging techniques (Paas et al., 2003).

Extraneous cognitive load: The type of cognitive load that arises from sensory inputs that are not beneficial to learning (Leppink et al., 2013). Measured by subjective rating scales, physiological measures, performance measures, or neuroimaging techniques (Paas et al., 2003).

Germane cognitive load: The type of cognitive load that arises from the effort required to transfer new information from the working to the long-term memory and properly integrate it with the existing knowledge (Leppink et al., 2013). Measured by subjective rating scales, physiological measures, performance measures, or neuroimaging techniques (Paas et al., 2003).

Intrinsic cognitive load: The type of cognitive load that arises from the complexity of the subject learned and the lack of previous knowledge held by the learner (Leppink et al., 2013). Measured by subjective rating scales, physiological measures, performance measures, or neuroimaging techniques (Paas et al., 2003).

Traditional instruction: Instructional techniques that have been used for centuries in education. These methods typically involve a teacher delivering information to students through lectures, textbooks, and other resources. The teacher is seen as the primary source of knowledge, and students are expected to listen, take notes, and memorize information (Cosgrove & Olitsky, 2018).

VRLE: An immersive learning experience delivered via a head-mounted device that creates a 3D virtual teaching environment (Vesga et al., 2021). It enables learners to acquire knowledge through the senses of hearing, sight, and touch while interacting with the environment (Sweller, 1988).

Assumptions

This study was based on several assumptions. Secondary students taking a science course from the control group of no simulation and the group participating in simulation responded to a survey regarding their experience of the mental effort needed to learn the new material. The underlying assumption was that students could self-measure and be honest about their efforts when completing the survey. Standardized curricula and the required professional learning communities mitigate any issues with the assumption of similarity of courses between teachers; however, I assumed that teachers did what they claimed they would do, and therefore, the content was consistent regardless of the teacher; however, there was no direct observation of the instruction. Another assumption was that students were placed appropriately in a science course matching their prior knowledge and learning achievement.

Scope and Delimitations

In this study, I focused on high school students without regard for gender, learning disabilities, or other factors that could affect the mental effort of a student learning a new scientific concept. Moreover, the VRLE used in this study was limited to the science VR content created by one company, Veative, an immersive education technology group. The study focused on the cognitive load as measured from the student's perspective, and I did not consider any measurement of learning outcomes or knowledge retention. The scope of the study focused on the measurement of the three types of cognitive load separately and did not provide an overall cognitive load measurement. The boundaries of this study were limited by the population that was not included, such as students with severe learning disabilities, the school's location in an affluent suburban neighborhood, and students not between 15–18 years of age.

Limitations

Sweller's CLT is a widely used framework for understanding how learning occurs and how cognitive resources are allocated during learning. However, there are also some limitations to the theory that should be considered:

1. Limited applicability to real-world situations: Sweller's CLT is based on laboratory studies and may not be entirely applicable to real-world situations where learners are often confronted with complex and dynamic learning environments (Paas et al., 2004). I addressed this issue in the current study by using Leppink's MES, which relies on student perception and not external measures.
2. Focus on cognitive load rather than other factors: CLT emphasizes cognitive load as the primary factor affecting learning, but other factors, such as motivation, engagement, and affective factors, can also affect learning outcomes (Kirschner et al., 2011). My choice of population addressed this issue by working with students that are very similar in motivation and abilities.
3. The narrow view of instructional design: CLT is primarily focused on how instructional design can reduce cognitive load, but it may not fully consider other vital aspects of instructional design, such as learner motivation, interest, and engagement (Kirschner et al., 2011).
4. Limited attention to individual differences: CLT does not account for individual differences in learners' abilities, prior knowledge, learning styles, or motivation, which can also influence the effectiveness of instructional design (Clark & Mayer, 2016). This limitation was one of the boundaries of the current study.

Significance

The significance of a study can be judged by the potential contributions the study may make that advance knowledge in the discipline. The current study contributes to educational technology by providing further evidence to the question of the effectiveness of VRLEs as a mode of science instruction for secondary students. Providing empirical research on whether the cognitive load experienced by high school students who use VRLEs during science instruction differs significantly from those taught using traditional methods helps inform how the technology is used. The study results provide teachers and administrators with the information they need to make informed decisions about pedagogy and to help determine when/if using VRLEs benefits students during science instruction. Making research-based and informed decisions about instruction may increase students' academic success. Increased academic success is a powerful force in the efforts for social change. Ultimately, the goal of this study was to use the findings to help others make informed decisions about best practices for teaching science to high school students and whether VRLEs should be incorporated into science instruction. This decision could help improve student learning outcomes and ensure that science education is as effective as possible.

Summary

In Chapter 1, I presented the primary themes and overarching problems addressed in the study. The background section contained a summary of the existing literature that supported this study. The problem and purpose statements indicated the study's focus on the cognitive load experienced by high school students learning science in a VRLE compared to the traditional methods. The development of the research questions addressed the scores of the three types of cognitive load as measured by Leppink's MES. In the theoretical framework section, I described the use of Sweller's CLT to support the scope and nature of the study. In the section on the nature of the study, I explained the rationale behind utilizing a causal-comparative and nonequivalent control design using archival data. Additionally, definitions were provided to clarify key terms and terminology in the context of the study. The assumptions, scope and delimitations, and limitations section included a discussion of the parameters and constraints of the study as well as the measures I took to address the limitations. I concluded Chapter 1 by explaining the study's significance and potential impact on secondary science education. In Chapter 2, I will describe the strategies used for conducting the literature search, delve into the details of the theoretical framework, and provide a comprehensive review of the current literature related to the current study.

Chapter 2: Literature Review

In this quantitative study, the problem was that with the increased use of VRLEs in high school science classrooms, it is not yet fully understood what the effects are on students' cognitive load, which can hinder their learning if it is too high. I used archived data that measured the three types of cognitive load separately to see if learning via VRLEs lowers the intrinsic load of science education and does not increase the extraneous load, allowing for more of the germane load. The literature has shown that VRLEs increase student interest in science (Astuti et al., 2020; Garduño et al., 2021), but there is a lack of consensus regarding their cognitive benefit. Some researchers have found that VRLEs increased students' retention and recall scores (Baceviciute et al., 2021; Lee et al., 2010; Sun et al., 2019), while others found the opposite (Makransky, Terkildsen, et al., 2019; Parong & Mayer, 2021), and the same is true regarding cognitive load. This study adds to the literature by measuring the three types of cognitive load separately and as measured by students' perception of the difficulty encountered in a high school science lesson. The purpose was to determine if the cognitive load experienced by high school students who use a VRLE during science instruction is significantly different compared to those taught using traditional methods.

Chapter 2 starts with a description of my literary search strategy and follows with a discussion of the theoretical foundation. In the literature review, I look at studies that measured cognitive load during different uses of VR, not limited to VRLEs. I separated the research into those studies that focused on a sense of presence, cognitive effects, and the different ways to measure cognitive load. I then review the literature regarding how

VR is used in science education, first focusing on its effect on scientific attitude and then on its effect on academic performance. I then look at its use for virtual field trips and those that use it as a virtual lab. In the last section of the literature review, I explore research about the cognitive load of high school science education, dividing the literature into a discussion on the inherent complexity of the subjects taught and the reasoning skills taught simultaneously with the material. I then examine the different teaching methods and tools used to lower cognitive load.

Literature Search Strategy

The Walden University Library was the primary access point to the sources used in this study, including books, reports, and peer-reviewed and empirical research articles published between 2017 to 2023. The databases accessed included Education Source, Academic Search Complete, APA PsycInfo, Computers & Applied Sciences Complete, ERIC, and Teacher Reference Center. Specific themes were utilized to help narrow my research by using keywords that appeared during the research. The search terms used were *virtual reality or VR or augmented reality or virtual environment, cognitive load OR mental effort, high school or secondary school OR secondary education OR 9-12, science education OR science instruction OR science learning, science labs, abstract or complex, virtual field trips, cognitive load measurement*, and *load reduction instruction.* I used Mendeley software to save and organize literature under general themes. Microsoft OneNote was used to take notes on the literature and organize it according to problem statements and results as well as by theories and methodology. Saturation of the literature was reached when themes and conclusions kept reappearing with no new study findings.

Theoretical Foundation

This study's theoretical framework was based on Sweller's CLT. Sweller (2020) argued that when designing a learning experience, the workings of human cognition must be considered. For a learner to acquire new knowledge, information must travel from the external world into the working memory and from the working memory to the long-term memory (Sweller, 2016). The working memory is the processing center for the new information, and it is limited to dealing with four to seven elements at a time for about 20 seconds (Sweller, 2020). The working memory capacity increases proportionately to the amount of knowledge in the long-term memory so that a novice learner will have a much more limited capacity than an expert one (Sweller, 2016). Information enters the working memory through visual and auditory instruction, and if it is processed successfully, it then enters the long-term memory, where it can connect to previous knowledge (Sweller, 2016). Once connected to previous knowledge, it becomes part of a person's cognitive architecture, and the student can apply it to diverse situations (Mayer & Moreno, 2003). The amount of information held in the working memory at any given moment is called a cognitive load, and if information entering the working memory exceeds its capacity to process, it is called cognitive overload (Mayer & Moreno, 2003).

Types of Cognitive Load

CTL includes three types of cognitive load: intrinsic, extraneous, and germane (Sweller, 1988). Intrinsic load arises from the complexity of the subject learned and the lack of previous knowledge held by the learner (Leppink et al., 2013). A topic is more complex if it requires more elements interacting simultaneously in the working memory

(Sweller, 1988). For example, a multimedia presentation that used a complex sentence would require students to process more elements at once than a simple sentence. Sometimes multimedia can reduce the intrinsic load by using graphics or animation instead of words to explain a complex concept (Clark & Mayer, 2016). Most of the time, the intrinsic load of complex subject matters cannot be diminished except by increasing the learners' pretraining (Sweller, 2020). However, one of the promising aspects of VRLEs is that they could decrease the intrinsic load of complex concepts by allowing learners to observe the events as they happen instead of relying on static images and explanations (Hu Au & Lee, 2017).

The second type of cognitive load is the extraneous load, which stems from instruction elements that do not help the student process information, such as a lengthy explanation of a problem or an unrelated anecdote in a lecture (Sweller, 1988). Some presentations require students to hold more interacting elements in the working memory than the subject matter requires; in that case, the intrinsic load increases unnecessarily and, hence, is considered extraneous (Sweller, 2016). The extraneous load can also arise from environmental factors during instruction, like room temperature, noise from the hallway, children playing outside the window, poor lightning, and the like (Paas & Van Merriënboer, 1994). In multimedia instruction, the extraneous load can arise from using background music that is the theme from a famous movie or creating an illustration with a character associated with a famous video game or cartoon, all of which will bring a rush of unrelated content into the working memory (Clark & Mayer, 2016).

The third type of cognitive load is the germane load, which is the effort required to transfer new information from the working to long-term memory and properly integrate it with the existing knowledge (Leppink et al., 2013). Instructional features that aid the student in these cognitive processes of integrating and organizing new content decrease the germane load, where most mental effort should be directed (Leppink et al., 2013). The greater the intrinsic and extraneous loads of a task, the less room for the germane load before the learner experiences cognitive overload and nothing further is learned (Sweller et al., 2019). An efficient instructional tool lessens the intrinsic load, eliminates extraneous load, and focuses all mental effort on the germane load so deep learning can occur (Sweller, 2020). An example of an efficient tool is a PowerPoint presentation that segments new information into small, simple elements; avoids anything irrelevant to the topic; and adds worked samples or graphic organizers that help the creation of mental schemas in the student.

CLT Applied to Educational Technology

CLT is an educational theory that helps design instructional methods that optimize what is known about human cognition (Sweller, 1988). Sweller also developed the theory to determine the efficiency of existing instructional strategies (Paas & Van Merriënboer, 1994). In CLT, Sweller (2016) provided several instructional effects to reduce cognitive load, including the modality, split-attention, redundancy, coherence, and personal and engagement effects. Mayer and Moreno (2003) and later Clark and Mayer (2016) applied these cognitive load effects to design multimedia learning materials. Mayer and Moreno explained that working memory comprises a visual processor and a verbal processor, and each processor is limited in capacity and can be overloaded. In the modality effect, Clark and Mayer explained that instructional designers must balance information delivery between the visual and auditory channels by conveying words via audio instead of text and leaving the visual channel free to process images. This balancing between the two channels can also lessen the split-attention effect, in which related concepts are visually separated. If an instructional designer separates a picture and its explanation into two slides, the working memory must integrate them. There is a more significant cognitive effort in learning when the working memory holds the picture longer until it reaches the explanation and creates a single concept. It is better to use the visual channel for the picture and audio for the explanation and avoid the extra load on the working memory. However, suppose the information is very complex. In that case, it is better to use written text for the explanation in another slide rather than audio due to the transient nature of the spoken word. Sweller explained that spoken cannot be retained in the working memory for an extended time since it is continually replaced by more audio; however, as the redundancy effect explains, the instructional designer should not repeat the same information through the visual and auditory channels. The repetition would cause the working memory to use twice the processing power to arrive at the same knowledge (Clark & Mayer, 2016). In addition, instructional designers need to be aware of the coherence effect, which is the need to keep all information presented on the topic. Clark and Mayer warned of the danger of adding noninstructional material, such as background music or unrelated graphics, to a lesson requiring additional audio or visual processing by the working memory and increasing the extraneous cognitive load. At the same time, the use of extraneous material could be justified if it increases psychological engagement between the learner and the content, which in turn decreases the germane load for the student, as explained in the engagement and personalization effects (Clark & Mayer, 2016).

Since its inception, CLT has been used in various educational technology research to help instructional designers, teachers, curriculum developers, and administrators better understand how media helps or hinders students' learning. CLT was used in the early 1990s to study student learning in hypertext environments (Cates, 1992; Laurel et al., 1990) and continued to be used as multimedia as an instructional tool became more complex. In the 2000s, CLT was used to determine which instructional approach was more effective at content delivery and produced higher performance scores. For example, a study compared three types of software to teach algebra (May, 2005). Another one compared three computer-based chemical engineering design concepts (Aubteen Darabi et al., 2007), and a third compared a lecture method and a web-based method of instruction (Chilton & Gurung, 2008). CLT continued to be used in the 2010s to evaluate educational technologies. For example, it was used to show how an online project-based approach caused cognitive overload compared to the classroom approach (Chen, 2016), which type of lecture model best focuses student attention (Hadie et al., 2016), and how detailed illustrations benefit or distract learning (Yu et al., 2017). Lastly, in the last decade, CLT has been used in several AR and VR educational technology studies. For example, CLT was used to research the cognitive load levels experienced by secondary students learning to speak English using an AR application (Küçük et al., 2014). It was also used to determine the effectiveness of desktop VR software for high school biology students with differing learning styles (Lee & Wong, 2014). A third example studied the relationship between cognitive load in a VRLE and learners' perceived sense of presence (Huang et al., 2020).

Rationale for the Use of the CLT

The cognitive effects of CLT as applied to multimedia design provide a rationale for why a VR world could potentially cause cognitive overload (Clark & Mayer, 2016; Mayer & Moreno, 2003). The psychological engagement gained by a completely captivating environment such as a VRLE could be too much for the brain to process, increasing extraneous cognitive load. Moreover, in a VRLE, learners need to move around and click on objects, and this embodied learning adds a new sensory input to the working memory (Skulmowski & Rey, 2017), and there is a need to find out how this affects learning. Many authors have agreed that measuring the cognitive load of an instructional instrument is the best way to judge its efficiency (DeLeeuw & Mayer, 2008; Leppink et al., 2013; Paas et al., 2003; Skulmowski & Rey, 2017; Sweller, 2020). An instructional instrument is considered more efficient if higher performance scores are obtained than expected, given the amount of mental effort (Ayres, 2006). Cognitive load theorists rate instructional strategies according to an efficiency score, which is a combination of subjective cognitive load and performance measures (Clark & Mayer, 2016; DeLeeuw & Mayer, 2008; Mayer & Moreno, 2003). Based on the literature, it was appropriate to use CLT to compare science instruction using VRLEs with science instruction using a traditional method to see which is more efficient.

Measurement of Cognitive Load

The measurement of cognitive load is a vast topic in the literature. I have organized the extant literature into two types of measurements. The first is the objective methods that are either indirect or direct. The second is the subjective methods comprising the multiple questionnaires and surveys developed for subjects to identify their perceived cognitive load.

Objective Measurement

At the inception of the CLT, Sweller (2020) did not measure cognitive load directly instead relying on measuring learning outcomes and error counting to discern the presence of cognitive overload. Suppose students performed better in a posttest after one type of instructional design than those students taught with a different design. In that case, it is reasonable to conclude that the former design produced less cognitive load. Performance is still the most common method for measuring cognitive load because it is objective and straightforward; however, it cannot measure each type of cognitive load separately (Sweller, Merrienboer & Paas, 2019). A more direct method of measuring cognitive load is to measure physiological changes like heart rate variability (Paas et al., 1994), pupil dilation (Van Gerven et al., 2004), or using electroencephalography (Antonenko et al., 2010; Montgomery et al., 1995). This type of measuring requires special equipment and can be cumbersome for the students (Morrison et al., 2014). Another direct measurement method is the dual-task approach, where the timing of a secondary task, such as foot-tapping, provides continuous measurement of cognitive load (DeLeeuw & Mayer, 2008; Park & Brünken, 2015). This method can continuously measure cognitive load without cumbersome equipment; still, like all the other objective methods, it cannot discern between the three types of cognitive load.

Paas and Van Merrienboer (1994) concluded that the best method to measure cognitive load was a subjective measure of mental effort, which is the amount of working memory used by the learner to perform a task. They developed a scale for students to self-report their mental effort after a lesson. They obtained similar results when measuring cognitive load physiologically or with their self-reporting scale. The scale is an accepted measurement of self-reported cognitive load (Ayres, 2006; Sweller, 2011); however, this scale does not distinguish between the three types of cognitive load either, and this would become the most sought-after goal in the field (Ayres, 2006; Krell, 2015; Morrison et al., 2014).

Several other self-report surveys have been developed to measure the three types of cognitive load separately. The Naïve Rating Questionnaire is reliable, valid, and has been used successfully (Klepsch et al., 2017). However, it was inappropriate for the current study because it has never been used with a similar population. Several studies have used the National Aeronautics and Space Administration Task Load Index to measure cognitive load. However, its validity as a measure of the total cognitive load has been questioned since it does not produce results like the Pass scale, and its use is limited to a valid measure of intrinsic load (Morrison et al., 2014; Naismith et al., 2015).

I chose to use Leppink et al.'s (2013) MES, one of the subjective measurements available. This self-report survey was the first reliable scale to measure the three types of cognitive load separately. Leppink et al. examined existing scales that measure either total cognitive load or one type of cognitive load and noticed the wording of the questions and type of measurement scale. Cognitive load has to be measured at an interval level, but some of the existing scales failed to do so. There was a lack of consistency on the type of scale used, some had a range of 0–100, while others simply 0–3 and others did not use numbers but simply low-medium-high choices (Leppink et al., 2013). The MES has questions worded to measure the different types of cognitive load, but the same 9-point Likert scale for all of them (Leppink et al., 2013). The MES showed internal consistency with reliability scores that revealed Cronbach's alpha values that matched expectations; likewise, the validity scores and the factor analysis matched the expected scores (Leppink et al., 2013). Additional reliability and validity scores were performed with a population of 232 high school students (Cook et al., 2017) and again after changing the wording to reflect a computer science course instead of a statistics course (Morrison et al., 2014) and both obtained similar results to the original. Leppink et al.'s MES has been used in multiple other studies to measure the cognitive load of handheld devices by secondary students (Becker et al., 2020), to determine which lecture method produced lesser cognitive load (Hadie et al., 2016), and to measure the cognitive load of mobile AR system (Ibili & Billinghurst, 2019). For these reasons, Leppink et al.'s MES was a good choice for the current study because the CLT aligned with this instrument and helped me interpret the results because the theory and instrument also aligned well with the study's purpose and research questions.

VR and Cognitive Load

VRLEs are defined as 3D worlds that are computer created, immersive and interactive, and accessed via a head-mounted display (Parong & Mayer, 2018). The immersive nature of a VRLE makes it different from desktop VR or AR content and necessitates studies that focus on how this immersion affects learning (Makransky, Terkildsen, et al., 2019). The literature surrounding desktop VR is consistent on its efficacy for learning, with numerous studies showing that it did not increase the cognitive load of students (Chen, 2006; E. Lee & Wong, 2014; Makransky, Mayer, et al., 2019; Parong & Mayer, 2018; Vesga et al., 2021). However, the elements needed to make a VRLE world realistic, including high-resolution graphics and extensive details, may not be necessary for learning and could cause extraneous cognitive load (Frazier et al., 2021; Makransky, Terkildsen, et al., 2019; Zhao et al., 2020). On the other hand, the VRLE's ability to mimic the real world creates the learner's sense of presence and increased engagement (Huang et al., 2020), which could lead to deeper learning. It is also possible that VRLE could positively affect learning things that require manipulation of the environment, learning procedures, or task performance (Shin & Park, 2019). In contrast, it could harm other types of learning, such as abstract concepts and ideas (Huang et al., 2020). It is also possible that VRLE's learning efficacy depends on a student's learning style (Huang et al., 2020), students' spatial abilities (Lee & Wong, 2014; Sun et al., 2019), or on a student's gender (Ariali & Zinn, 2021; Ibili & Billinghurst, 2019).

Sense of Presence

The technology used to create VRLEs creates a fully immersive environment that can translate into the feeling of being there, called the sense of presence (Parong & Mayer, 2021). Study results are divided on whether or not the sense of presence creates only emotional or cognitive engagement. According to Dewey's theory of interest, when students are more present, they learn more (Jonas, 2011). In an older study, desktop VR results showed that a sense of presence did correlate with increased learning outcomes (Lee et al., 2010). However, a study of 52 university students comparing VRLE and desktop VR showed that knowledge gained by the desktop VR group was significantly greater with a p-value of .006, even though the sense of presence was lower (Makransky, Terkildsen, et al., 2019).

Similarly, a study of 61 adults comparing VRLE to a PowerPoint slideshow found the VRLE group to have significantly more presence. Still, the slideshow group scored significantly better on knowledge and transfer tests, with a p value of less than .001 (Parong & Mayer, 2021). In other studies, the increased sense of presence found in VRLE did yield better learning outcomes. For example, Sun et al. (2019) examined whether a VRLE group performed better on knowledge retention tests than a slideshow group. Another case is a study that compared reading a medical pamphlet in a classroom with reading the exact text in a VRLE of a medical office, with the latter group scoring better in retention tests (Baceviciute et al., 2021). Given the mixed results, there is a need for further examination to understand better how VRLE can contribute to learning beyond just the sense of presence.

Cognitive Effects

Meyer et al. (2019) examined whether the cognitive effects of CLT apply to VRLE or if they apply differently in differing media. Multiple authors have attempted to answer this question by performing studies focusing on one cognitive effect. An example of this is the split-attention effect that must be balanced with audio's transient nature in animations and videos. Studies have shown that adding static images to animations lessens the cognitive load and improves test scores by providing visual cues that direct the learner's attention and help retain the information (Shin & Park, 2019). When a study aimed to replicate these results by adding static images inside a VRLE the results showed a decrease in cognitive load compared to the same VRLE without the images. Still, no significant difference was found in the test scores (Shin & Park, 2019). Another study also focusing on adding visual signaling to VRLE to reduce cognitive load found no significant difference between the cognitive load of the group in a VRLE with visual signaling and one without it (Nelson et al., 2016). The mixed results indicate the need to study cognitive load in VRLE further.

Another cognitive effect is segmenting information into small parts to lessen element interactivity, decreasing the intrinsic load (Clark & Mayer, 2016). Parong and Mayer (2018) aimed to lessen the cognitive load of VRLE by applying the segmentation effect and asking students to pause and write summaries before continuing through the lesson. The VRLE group with segmentation scored better in posttests while maintaining the same level of interest (Parong & Mayer, 2018) showing the cognitive effect to work in VRLE just as it does in other media.

Zhao et al. (2020) advocated for using cognitive effects in VRLEs to ensure students are cognitively active and not just passively engaged. Their study tested the summarizing strategy by randomly separating 75 college biology students into two VRLEs groups and two interactive video groups. The summarizing strategy was equally beneficial in knowledge retention and lower cognitive load in the VRLE group with $p < 0.01$ and the interactive group with $p < 0.04$ (Zhao et al., 2020). These results show that once again, the cognitive effect works similarly in a VRLE as in other media.

Measuring Types of Cognitive Load

One way to further explore the mixed results in the literature is to measure the three types of cognitive load separately when using VRLEs. This approach might shed light as to why in some studies, the cognitive load of VRLEs was higher than the non-VRLE group (Frederiksen et al., 2020; Makransky, Terkildsen, et al., 2019; Parong & Mayer, 2018; Vesga et al., 2021) while in others the cognitive load of the VRLE group was lower (Baceviciute et al., 2021; Dan & Reiner, 2017; Sun et al., 2019). Moreover, multiple studies found no difference in cognitive load between the two groups (Frederiksen et al., 2020; Huang et al., 2020; Lamb et al., 2018; Luong et al., 2019; Nelson et al., 2016; Shin & Park, 2019; Zhao et al., 2020). The effectiveness of an instructional tool is not simply a lowering of overall cognitive load but to lower the intrinsic and extraneous load and to increase the germane load while avoiding cognitive overload (Sweller, 2020). A few of the studies listed above attempted to measure the types of cognitive load separately. For example, Parong and Mayer (2021) concluded that the VRLE self-reported scores showed a significant difference (p =.004) in extraneous loads but not intrinsic or germane loads. They also found no difference in the workload as measured by EEG (Parong & Mayer, 2021). Vesga et al. (2021) showed that the increased presence of VRLEs led the student to exert more effort which led to high GL readings but did not find evidence of cognitive overload. There is need for further studies that measure cognitive load separately for students engaged in VRLEs.

VR and Science Education

VRLEs use in the high school science classroom has dramatically increased in recent years and is expected to continue (Zhang et al., 2020). Many benefits have been associated with VRLEs and their use in secondary science education. Among them is an improved scientific attitude (Astuti et al., 2020; Cheng & Tsai, 2019; Hu Au & Lee, 2017; H.-M. Huang & Liaw, 2018; Makransky & Lilleholt, 2018). The increased performance by students in posttests and knowledge retention tests (Bogusevschi et al., 2018; Cai et al., 2017; Hatchard et al., 2019; Makransky, Terkildsen, et al., 2019) is another benefit, as well as the possibility and benefits of virtual field trips to locations otherwise unreachable by the students (Cheng & Tsai, 2019; Lin et al., 2011; Petersen et al., 2020). Moreover, the possibility of online science courses or advanced lab sciences in rural areas due to the expediency of virtual labs is another area of growth (Bogusevschi et al., 2020; Hatchard et al., 2019; Makransky, Terkildsen, et al., 2019; Torres et al., 2015). In this section, I organized the literature by VRLEs and scientific attitude, student performance, virtual field trips, and virtual labs.

Scientific Attitude

One way that VRLEs has been examined in high school science classrooms is their effect on students' attitudes toward science. There is a growing decline in interest in advanced science classes at the secondary and post-secondary levels (Bogusevschi et al., 2020). The abstract nature of scientific concepts can add to students' belief that science is too difficult or tedious. Several studies have shown that VRLEs can increase students' interest in the sciences and their attitude toward science (Astuti et al., 2020; Garduño et al., 2021). Science attitude is vital because when student interest in a subject increase, learning achievements also tend to increase (Boda & Brown, 2020). In a study of 96 high school students comparing a chemistry lesson taught in a VRLE versus one taught in a conventional classroom, 95% showed increased scientific attitude and critical thinking skills (Astuti et al., 2020). Similarly, in a study of 304 high school students using VRLEs to learn chemistry, 72% showed increased interest and attention (Garduño et al., 2021). Moreover, VRLEs showed an increased interest and attitude with a group of 308 undergrad students (Huang & Liaw, 2018), with 65 high students learning wave energy (Huynh et al., 2016), and with 16 high students learning about plastics and the environment (Keller et al., 2018). Numerous other studies have found an increase in interest and attitude (W. K. Liou & Chang, 2018; Makransky & Lilleholt, 2018; Nersesian et al., 2019b; Parong & Mayer, 2021).

Using VRLEs in high school science classes affects an interest in science for several reasons. First, VRLEs help make tangible and present processes at the molecular level or inaccessible locations, making science more attractive to students (Nersesian et al., 2019b; Petkov et al., 2019). VRLEs also allows students to manipulate objects they are studying, taking advantage of the benefits of embodied learning (Boda & Brown, 2020). Scientific attitude is more than just an increase in interest; it is an attitude of the mind open to inquiry and problem-solving (Astuti et al., 2020).

Performance

The increased interest in science and a desire for inquiry raised by the use of VRLEs in high school science classes has also led students to score higher than the non VRLE students in posttests and retention knowledge tests (Bogusevschi et al., 2018; Jitmahantakul & Chenrai, 2019; Lai et al., 2022; W. K. Liou & Chang, 2018; Nersesian et al., 2019b; Southgate, 2019). Ninety-three students from three different high schools taking a geology class were given a lesson using a VRLE (Jitmahantakul & Chenrai, 2019). The students took a pretest and a posttest, and the results showed that the average posttest scores were significantly higher than the pretest scores with $p = .05$, and showing learning gains of the students improved by 22%-28% (Jitmahantakul & Chenrai, 2019). Similarly, 52 secondary students were divided into VR and non-VR groups and learned about the water cycle, and 74% of the VR students showed knowledge gain between the pretest and posttest compared to 48% of the non-VR (Bogusevschi et al., 2018). Positive test results have also been shown with elementary students, as in the pretest and posttest study of 24 fifth graders, student learning improved by 27.67% when learning about ecosystems (Khotimah et al., 2021). More important than just increased test scores, in a study with 56 Grade 9 students learning science, the VR group showed higher order cognitive abilities such as critical thinking and problem-solving (Southgate, 2019).

However, research on VRLE and performance has also been done with older students. For example, in a study with 54 university students, the VR group did not show more significant learning outcomes than a traditionally taught group (Makransky, Terkildsen, et al., 2019). Results showed that the non-VR group learned significantly more knowledge with $p = .040$. Similarly, in another study of 55 college students studying biology, the non-VR group had significant learning gains with $p = .003$ (Parong & Mayer, 2018). Current research shows that VRLE tends to affect the younger populations differently than the older ones, indicating a need for studies like this one, which focuses on the effect VRLE has specifically for the high school population.

Virtual Field Trips

Another widespread use of VRLEs in science classrooms is virtual field trips. VRLEs allow students to experience and explore places and phenomena that are otherwise unreachable, dangerous, or too expensive to visit. A virtual field trip is a way to visit a location without traveling there and allows students to have a first-person view of the place (Lin et al., 2011). A group of high school students participated in a virtual field trip to Greenland to study the effects of climate change. The experience of seeing the effects of global warming on the ice sheet created a desire for action and further inquiry (Petersen et al., 2020). The students took a pretest and posttest regarding interest and knowledge of climate change, and there was a significant increase in interest with $p < .0005$ and an increase in transfer scores (Petersen et al., 2020). Climate change is just one example of how a complicated scientific topic can be witnessed or experienced inside a classroom or laboratory. Virtual field trips allow students to interact with and analyze a location in ways that authentic travel does not, enhancing the learning value of the virtual trip (Cheng & Tsai, 2019; Petersen et al., 2020). For example, a group of high school students studying geology who participated in a virtual field trip to a national park were asked to figure out why the water in the lake was bubbling. Since there are no physical limitations in a virtual field trip, students could go underwater and explore in a way not possible in a real trip to the same park. (Lin et al., 2011). In this virtual field trip, students could watch a volcanic eruption up close and in slow motion if desired.

Another aspect of the value of virtual field trips is their impact on lower socioeconomic schools or remote rural areas where a field trip to otherwise easily accessible locations is not possible (Kenna & Potter, 2018). Teachers use field trips half as often as they did 30 years ago (Kenna & Potter, 2018). Teachers cite a lack of funds, refusal to admit liability by schools, and lack of Americans with Disabilities Act accommodations, among the many reasons they no longer plan field trips. Therefore, virtual field trips provide an alternate way to create experiential learning opportunities not hindered by these concerns (Kenna & Potter, 2018). Virtual field trips also provided a way for students to have access to these opportunities during the global pandemic of COVID-19, not only during the remote learning of the shutdown but also afterward with the limitations imposed by legal restrictions (Gielstra et al., 2021; Mcpherson et al., 2021; Price & de Ruiters, 2021). International travel became difficult due to border closures, and domestic trips became almost impossible due to the restrictions that led to increased use of virtual field trips after the shutdown (Gielstra et al., 2021).

Moreover, the pandemic deepened the lack of funds available for actual field trips due to the higher cost of making face-to-face education safe for everyone (Gielstra et al., 2021). Ultimately, virtual field trips are not meant to replace actual field trips but to provide a way for students to experience some of the benefits of actual field trips when these are not possible (Seifan et al., 2020). In summary, field trips have been shown to contribute to the increase in scientific attitude and interest, making virtual field trips a valuable part of the science classroom in instances where it is not feasible to have an actual field trip.

Virtual Labs

Due to the high cost of obtaining and maintaining such equipment, many secondary schools lack the proper equipment in the science labs to run the experiments required in advanced science courses like chemistry and physics (Bogusevschi et al., 2020). Many schools do not have an introductory lab due to a lack of space and funds, making science harder to learn. Students cannot perform the activities to develop the cognitive processes necessary to understand the content (Torres et al., 2015). VR labs provide an affordable solution to these situations, making advanced science courses available to a larger part of the population (Bogusevschi et al., 2020). In the past, schools in these situations have turned to online courses to provide access to these advanced courses. However, the online courses lack a lab component, making the courses purely theoretical and hard to comprehend (Kumar et al., 2021). VR labs are one of the ways online schools can offer advanced science courses that universities now accept to fulfill a lab requirement.

VR labs allow students to repeat experiments as often as desired, see how the outcomes differ, and attempt further experiments. The students can manipulate the specimens, tools, and chemicals at their disposal since it is a virtual space with no additional cost other than time (Kumar et al., 2021; Lai et al., 2022). The increased availability of resources makes virtual labs an excellent tool for scientific inquiry and exploration, which increases comprehension (National Research Council, 2000). In a study of 66 high school students separated into VRLE and Desktop VR, the Desktop group scored higher in the immediate posttest; however, the VRLE group scored higher in the later follow-up test (Lai et al., 2022). Virtual labs are helping high school students to understand the concepts better because they can experiment with the material in ways beyond those prescribed in a textbook.

Cognitive Load and Science Education

Science education is an essential topic for governments worldwide because the only way to move forward and grow is not based on the number of natural resources but on innovative solutions (Uçar & Sungur, 2017). Economic progress for countries has often come about when they find ways to be more efficient, decrease pollution, cure more diseases, or develop better software; in other words, scientific development has been the key to moving them forward (Johnson, 2012; Murphy et al., 2018). The link between scientific progress and economic growth has led governments to be involved in improving science education. Wishing to increase the number of scientists in the field, governments are recruiting students who want to study science at the university level; however, that requires students who enjoy and understand science at the secondary or high school level (Aguilera & Perales-Palacios, 2020). Researchers worldwide report that high school students are losing interest in science and scoring poorly in national science scores. Low scores in science courses and tests are a global problem, with researchers reporting low scores in Nigeria (Achor et al., 2019; Etobro & Fabinu, 2017), Spain (Aguilera & Perales-Palacios, 2020), Turkey (Bal-Incebacak et al., 2019; Uçar & Sungur, 2017), Indonesia (Hidayati et al., 2020), Jamaica (Lodge, 2021), Bhutan (Nidup et al., 2021), Netherlands (Santi & Gorghiu, 2019), Norway (Steier & Kersting, 2019), Taiwan (Weng et al., 2018), and the United States (Murphy et al., 2018; National Center for Education Statistics, 2015; Patall et al., 2019; Zhang, 2019). When science scores and student interest are reported low globally, the problem is likely associated with something inherent to the subject. The problem cannot be attributed to local circumstances like teaching methods, textbooks, or classroom management. When concepts are complex, they have a high intrinsic cognitive load, which can easily lead to cognitive overload if it exceeds the working memory's capacity (Sweller et al., 2019). Suppose the student becomes overloaded, as it might with learning science content. In that case, students cannot move the information into the long-term memory, which is joined with prior knowledge, and, therefore, cannot be understood (Makransky, Terkildsen, et al., 2019).

The literature on cognitive load and science education has been divided into three parts. First, research that focuses on the subject matter's inherent difficulty, then on the reasoning schemas needed to learn the subject matter, and lastly, literature on interventions to decrease the intrinsic load of the subject matter or eliminate the extrinsic load as much as possible.

Inherent Complexity of Learning Science

Science education encompasses different disciplines, and research studies have shown that learning science is inherently complex. Students and teachers of secondary-level physics agree that learning about Newton's Laws, waves, energy quantization, and others is difficult to comprehend (Achor et al., 2019). Concepts of space and time and Einstein's theory of relativity are also challenging for students to grasp because they contradict students' reality experiences (Steier & Kersting, 2019). Researchers also found Faraday's law of electromagnetic flux is a tricky concept to teach at the secondary level due to its lack of concreteness (Kokkonen et al., 2022). The complexity of these science topics poses a high level of intrinsic load that, despite different teaching methods, remains an obstacle to student comprehension (Achor et al., 2019; Aguilera & Perales-Palacios, 2020; Etobro & Fabinu, 2017; Hidayati et al., 2020; Santi & Gorghiu, 2019).

Science education comprises many topics that cannot be seen by the naked eye and of which students have very little prior knowledge. For example, in a study of 100 high school students, 45% rated organic chemistry as a complicated subject, and 59% preferred other aspects of chemistry (Nartey & Hanson, 2021). This same group of students named "preparation and chemical reactions of alkenes, preparation and chemical reactions of alkynes, structure, and stability of benzene, reactions of benzene, comparison of benzene and alkenes" as the most challenging things to learn (Nartey & Hanson, 2021, p. 331). These topics are abstract and difficult to observe with the naked eye, making teaching them as tricky as learning them. Survey results from 90 high school students showed chemistry as their least favorite subject even though they report to like their teachers (Nidup et al., 2021). Some areas of chemistry, like some of the physics curriculum, are not learned by natural observation but by reading textbooks and listening to explanations. Concerning the biology curriculum, a study of 400 students found that they had difficulty with topics of “nutrient cycling in nature, ecological management, conservation of natural resources, pests and diseases of crops as well as reproductive system in plants” (Etobro & Fabinu, 2017, p. 139). When students find science lessons more difficult than others, they start to believe that they cannot succeed, leading to disengagement with the subject matter (Patall et al., 2018). Less difficult ones must surround complex topics, so students do not get discouraged and stop trying. Using appropriate language and breaking down the information into small chunks is vital. Yet, a study of 450 students found that biology textbooks used complicated and confusing vocabulary and complex sentence structures (Lodge, 2021). The textbook language's complexity increases the lessons' intrinsic load and leads students to disengage with the content.

As lack of engagement decreases aptitude, the opposite is also true. When science lessons are presented engagingly, the students are more likely to increase their effort when challenged by a complex topic (Martin et al., 2020). Researchers have shown that keeping students engaged with the content has increased achievement scores, motivating students to spend more cognitive resources on complex topics and lowering the overall cognitive load (Martin et al., 2021). Cognitive overload can be avoided by decreasing the load of the material or by increasing the students' cognitive capacity, which can be achieved by increasing interest. Research also shows that students with more prior knowledge have a lower cognitive load which allows them to invest more cognitive resources to learn complex topics and are more likely to stay engaged. The study showed that prior knowledge significantly affected engagement with $p < 0.01$ (Dong et al., 2020). Educators have tried many ways to keep students engaged with science instruction, but engaging students experiencing high cognitive load levels is not easy.

Scientific Reasoning and Cognitive Load

According to CLT, information passes from the working memory to the long-term memory using cognitive schemas by which the information can be organized and united with prior knowledge (Sweller et al., 2019). Some schemas are formed simultaneously as primary knowledge is formed by intuition and other biological methods; other schemas are developed as a child grows and learns secondary knowledge (Sweller, 2016). The more complex the information being learned, the more advanced schemas the brain needs to integrate into long-term memory. Science education encompasses complex concepts, and the adolescent brain does not always possess the necessary schemas to process that information (Wei et al., 2021; Zhang, 2019). Science teachers must teach the concepts and schemas, or reasoning capabilities needed to process the information correctly (Johnson, 2012).

Secondary science education provides the adolescent brain with new cognitive skills, such as problem-solving, scientific reasoning, and risk-taking (Bal-Incebacak et al., 2019). These skills allow students to make inferences about a problem, make hypotheses, and rearrange possible solutions (Bal-Incebacak et al., 2019), which can be used to process new information. Science education focuses on acquiring scientific notions and the critical skills necessary to research new ideas and further inquiry (Santi & Gorghiu, 2019). The Next Generation Science Standards drafted in 2013 moved away from listing scientific information to focusing on 21st-century skills that prepared students to discern the quality of scientific information in the modern world (Murphy et al., 2018; Wei et al., 2021). The modern world requires all citizens to understand and judge the technologies and advances an advanced society produces. The Next Generation Science Standards advocates the need for students to engage in scientific dialogue by identifying evidentiary data and making empirically informed decisions (Settlage & Southerland, 2019). The role of the secondary science teacher is to help students develop thought processes that enable them to sift through the scientific information they need to master (Settlage & Southerland, 2019). Developing reasoning skills and other epistemic tools adds to the intrinsic load of science education.

Interventions to Reduce Cognitive Load

Several studies address the problem of the high cognitive load of science education by introducing innovative tools or methods to teach science and then comparing the cognitive load before and after. Sometimes, the researchers did not introduce anything new but compared the cognitive load of the already used methods. For example, Chen et al. (2019) measured the cognitive load of having students draw diagrams after a lesson versus having them write a summary. The results showed that drawing diagrams lowered the cognitive load of the lesson more than summaries but that students with less prior knowledge needed assistance in how to draw them (Chen et al., 2019). It is important to note that even though prior knowledge still played a role in the amount of cognitive load experienced by the students, the diagram was still more helpful than the summary in reducing the total load experienced. Lardi and Leopold (2022) examined the diagram strategy more closely and compared having students create the diagrams themselves versus having students and teachers create the diagram together. The teacher-generated diagram was created interactively with the whole class so that the teacher followed the students' directions on what to draw and where; this strategy lowered the cognitive load more than the student-generated drawing strategy (Lardi & Leopold, 2022). Similar studies focused on other strategies, such as the use of worked examples (Saw, 2017), the use of argumentation style (Yildirir, 2020), and hands-on inquiry-based strategies (Zhang, 2019).

Another group of studies focuses on technological interventions that attempt to lower cognitive load. Like the previous studies, these compare the cognitive load of a group that uses a tool that the other group does not have. For example, Becker et al. (2020) studied the use of tablets to teach motion in a group of 286 high school students and reported that the group using the tablets had a lower extrinsic load with $p < 10^{-3}$ and an increased germane load with $p = 0.041$. It is important to note that an instructional tool is considered efficient if it lowers the intrinsic or extraneous load and increases the germane load (Sweller, 2020). Not all studies concluded the tool under examination was better, as was the case with interactive textbooks. Weng et al. (2018) measured the cognitive load associated with interactive versus paper textbooks with 44 junior high school students. They found that the group using the interactive textbooks had a higher perceived cognitive load as measured by the Leppink survey. The abundance of studies shows that science is inherently challenging to teach and learn. Examining innovations that could reduce secondary students' cognitive load could help improve students' understanding of science.

Summary and Conclusions

In this chapter, I discussed a consensus that virtual reality raises student interest in science (see Astuti et al., 2020; Garduño et al., 2021). VRLEs have been shown to help make abstract scientific concepts visible to the students and take them to remote locations, keeping students engaged and wanting to learn more (Nersesian et al., 2019a; Petkov et al., 2019). Yet there is disagreement about whether this new interest also leads to an increased scientific aptitude. Some authors have shown that, indeed, because of the increased interest, students put more effort into comprehending the new material, which often leads to better scores (Bogusevschi et al., 2020; Jitmahantakul & Chenrai, 2019; Lai et al., 2022; W. K. Liou & Chang, 2018; Nersesian et al., 2019a; Southgate, 2019). However, other studies show no cognitive gain in VRLE groups compared to other learning methods (Meyer et al., 2019; Parong & Mayer, 2018). The increased interest could be an emotional experience that can cause cognitive overload and not lead to cognitive gains (Makransky, Terkildsen, et al., 2019).

There is also agreement in the literature that students in secondary science education are often disengaged, frustrated, and do not pursue more advanced courses (Achor et al., 2019; Aguilera & Perales-Palacios, 2020; Bal-Incebacak et al., 2019; Etobro & Fabinu, 2017; Hidayati et al., 2020; Lodge, 2021; Nidup et al., 2021; Santi & Gorghiu, 2019; Uçar & Sungur, 2017).

Many technological instruments have been tried to improve the learning outcome results (Hochberg et al., 2020; Weng et al., 2018) with varying degrees of success. In recent years, VRLEs have also been used in science education to increase student performance, but there is disagreement as to its effectiveness. Those that oppose it claim that VRLEs use unnecessary instructional elements that cause the students to be distracted and experience cognitive overload (Makransky, Terkildsen, et al., 2019; Parong & Mayer, 2018). The ones that are in favor claim that VRLEs reduce the intrinsic cognitive load inherent to the subject matter by increasing student effort and by making things visible in a three-dimensional manner (Dan & Renier, 2017; Martin et al., 2021; Dong et al., 2020; Nersesian et al., 2019b; Petkov et al., 2019). It is uncertain whether the gain in decreased intrinsic load makes up for the increased external load. The gap that remains is studies that focus on measuring the three types of cognitive load separately and see if students experience cognitive overload. In Chapter 3 I describe my study methodology to explain how I collected the student data and which statistical tests I performed to analyze the results.

Chapter 3: Research Method

The purpose of this quantitative quasi-experimental study using a nonequivalent control group design was to determine if the cognitive load experienced by high school students who use a VRLE during science instruction is significantly different compared to those taught using traditional methods. To accomplish this, I used archival data from high school science students in various courses offered at a private high school in a western U.S. state that is referred to with the pseudonym of Private High School (PHS). The data points were derived from students who completed Leppink's MES after a science lesson; some were taught using a VRLE, and other students were not, thus, creating the opportunity for new comparison research. Outcome measures for the study were part of PHS's MES requested by the science department chair to be given to science class students after the introduction of VRLEs in the department. I conducted this study to address the gaps in the literature noted in Chapter 2 by measuring self-assessments of the students' three types of cognitive load.

In Chapter 3, I present the research methodology used in the study. I start by describing the research design and the rationale for implementing this research design in the study.. The methodology is discussed in the next section, including the study population, data collection, instrumentation, and outcome measures. The chapter concludes with threats to validity and the ethical procedures followed in this study.

Research Design and Rationale

The research questions were focused on the cognitive load experienced by high school students in Grades 9–12 who use VRLEs in science classes compared to those taught using traditional methods as measured by the self-reported MES. In this quantitative quasi-experimental design study with nonequivalent groups, I examined how the independent variable of method of science instruction affects the dependent variables of intrinsic cognitive load, extraneous cognitive load, and germane cognitive load. The research questions and corresponding hypotheses are presented in Table 1.

Table 1

Variables and Statistical Treatment by Hypothesis

H	IV	DV	Statistical treatment
When measured by Mental Effort Survey, the intrinsic cognitive load experienced by high school students who use VRLE during science instruction is significantly different than the cognitive load experienced by high school students receiving traditional science instruction.	Method of science unit instruction.	Sum of Mental Effort Survey scores for questions 1–3.	Compare the means of the two groups measured by Leppink's Mental Effort Survey with two independent sample *t* tests.
When measured by Mental Effort Survey, the extraneous cognitive load experienced by high school students who use VRLE during science instruction is significantly different than the cognitive load experienced by high school students receiving traditional science instruction.	Method of science unit instruction.	Sum of Mental Effort Survey scores for questions 4–6.	Compare the means of the two groups measured by Leppink's Mental Effort Survey with two independent sample *t* tests.
When measured by Mental Effort Survey, the germane cognitive load experienced by high school students who use VRLE during science instruction is significantly different than the cognitive load experienced by high school students receiving traditional science instruction.	Method of science unit instruction.	Sum of Mental Effort Survey scores for questions 7–10.	Compare the means of the two groups measured by Leppink's Mental Effort Survey with two independent sample *t* tests.

Note. H = hypothesis, IV = independent variable, and DV = dependent variable.

The quasi-experimental research with a nonequivalent control group design is similar to a proper experimental design in that it involves the manipulation of an independent variable to see its effect on a dependent variable; however, it differs in not using random assignment of participants to conditions (Burkholder et al., 2016). One example of a quasi-experimental design is a nonequivalent control group design, in which two groups are compared; still, the groups are not randomly assigned and may differ in other ways besides the intervention being studied (Coleman, 2022)

The design was appropriate to answer the research questions because students were enrolled and placed into previous existing groups comprised of several versions of the same course taught by different teachers at various times of the day. Placement into one group or another was somewhat random because teachers chose which method to use (i.e., VRLE or traditional). Scheduling the students with one teacher or another was based on availability and conflict with other required courses. The use of nonequivalent control groups is often seen in education due to the internal organization of schools (Coleman, 2022). Warner (2013) stated that although quasi-experimental designs provide weaker evidence, they have more substantial external validity because the interventions occur in a real-world setting such as schools.

Quantitative quasi-experimental designs can be used to study the effectiveness of educational technology interventions in various settings. In the current study, I used a quasi-experimental design to determine the impact of VRLEs on student cognitive processes to see if it is an effective tool to use in the classroom. Depending on the method the teacher of the course used, I assigned students' archival data sets to the control group or the group participating in the simulation with a VRLE. The research questions have three outcome variables: one independent variable with two categories. The three outcome variables are the subscores Leppink MES that focus on the different types of cognitive load and are measured using an interval/ratio 1–10 Likert scale.

The independent variable method of science instruction consists of two levels: One level is science students taught using a VRLE and the other level is students taught by traditional methods. The choice of VRLE was entirely up to the teacher, and the survey was completed during class time as required by the department chair. Teachers of the same course and level must meet weekly at PHS to ensure the same pace, content, and exams are used regardless of who teaches a course; however, the manner of teaching the content is up to the teacher's preference.

Methodology

In this section, I provide information regarding the population of the study, explain how archived data were gathered and how data included a traditional teaching method group and a group participating in a VRLE, present an overview of the instrument used, and describe how variables in the study were operationalized and analyzed. My role as the researcher was consistent with my present position as the educational technology specialist at PHS. A donation of 24 VR headsets was given to PHS, and they were preloaded with science lessons that match the national curricula. The teachers at the school were given the choice of using the headsets by requesting them from the IT department. The science department chair wanted to measure the effectiveness of the headsets before deciding to purchase more of them, so after consulting with me, the educational technology specialist, the department chair used the Leppink MES to measure the headset's effect. Archival data of the survey results were used in this study from the students enrolled in a high school science course.

Population

The target population for this study was high school students enrolled in a science class while attending PHS. The science courses at PHS are open to all students in Grade Levels 9–12 who meet the prerequisites. PHS is a private institution in the western U.S. focusing on college preparatory courses. The student demographics at PHS are 62% European American, 12% multiracial, 11% Hispanic, 9% Asian, 3% Other, 2% African American, and 1% Native American. The male-to-female ratio is 54% male and 46% female. Although the school is in a high socio-economic area, 30% of the student body received financial aid to attend the school. The archival data that I used has 223 students who completed the survey right after a science lesson during the 2021–2022 school year. The students in this population took a science class in which the teacher chose to use a VRLE as a method of instruction for one lesson or chose a traditional method. Students were assigned to one teacher or another by scheduling software based on schedule constraints.

Procedures for Using Archival Data

Researchers using archival data formulate a research question that fits the data accessed (Elder et al., 1993). Using data that fits a question is considered a type of purposive sampling. Nonprobability purposive sampling, also known as judgment sampling or subjective sampling, involves selecting a sample based on the researcher's judgment about which population members are most relevant or informative to the research question (Daniel, 2012). This method is often used when the researcher has a specific focus or theory in mind and wants to select a sample that will be particularly useful for testing or exploring that theory (Daniel, 2012). Nonprobability purposive sampling from archival data is a quick and efficient way to gather data, especially if the researcher has a clear idea of who or what they are looking for (Daniel, 2012). For the current study, the existing MESs were an appropriate choice for the study of the effects of VRLEs on student cognitive processes.

Power Analysis for Sample Size

Power is a statistical concept that refers to the probability that a study or experiment will detect an effect or relationship of a particular size, given a specific sample size and significance level (Frankfort-Nachmias & Leon-Guerrero, 2018). Power analysis for sample size in an archival data set is calculated differently since the sample size is predetermined (Cohen, 1969). One of the benefits of using archival data for power analysis is that it can provide a more accurate estimate of power because it is based on actual data rather than assumptions about the population or the effect size; therefore, in studies with archival data with a known sample size, the power, and calculation are post-hoc analyses that compute the power given an alpha, a sample size, and an effect size (Cohen, 1969). In this study, the total number of participants was 223, with 141 in the traditional method group and 82 in the VRLE group, and the effect size was set to the standard 0.5 and alpha to 0.05, which resulted in a powerful effect of 0.94. This is well over the 0.8 threshold required, which corresponds to an 80% chance of detecting an effect if it truly exists (see Appendix A). Cohen (1969) provided calculations that necessitated 64 participants per group for a power of .80, a medium effect size, and an alpha of .05. The VRLE group was comprised of those students who were taught with a

VRLE for one lesson, and the traditional group was comprised of all other students who were taught that same exact lesson in a traditional manner.

Procedures for Using Archival Data

The science department at the study site chose a survey to provide data showing whether a VRLE helped students understand and retain science lessons better than the traditional methods. The Leppink's MES was used. Permission to use the survey from the author is found in Appendix B. The survey was put into Microsoft Office Forms, and the science chair distributed it to the science teachers and instructed them to give the students the survey for each VRLE lesson they taught. In addition, teachers teaching the same lesson without VRLE were told to also give the survey.

Since I was aware of the use of VR headsets and the survey, I searched the literature when it was time to conduct this study. I found a gap regarding the use of VRLEs and their effect on the cognitive load experienced by students, precisely the three types of loads measured separately and from the student experience. I discussed the use of the existing student level data with the school principal and was granted permission to use it. I obtained Walden University's Institutional Review Board (IRB) approval on May 5th, 2023.

Instrumentation, Validity, and Reliability

The MES used in this study was developed by Leppink et al. (2013) as a multi-item measurement of cognitive load to obtain a global composite measurement for each type of cognitive load. The instrument is composed of nine questions using a 10-point Likert scale. The authors used Sweller's (2011) CLT to describe cognitive load, especially the qualities of cognitive load and how to know if cognitive overload has occurred. In the instrument, 0 means *not at all the case* and 10 means *completely the case* as students are asked about the difficulty and complexity of a statistics lesson. Multi-item measurements are more reliable than single-item measurements and are measured for internal reliability (Warner, 2013). The MES has three questions to measure intrinsic load, three to measure extraneous load, and four to measure germane load. The validity of a measurement tool refers to its ability to accurately measure what it is intended to measure (Frankfort-Nachmias & Leon-Guerrero, 2018). In the case of the Leppink MES, the instrument's validity depends on its ability to assess the three types of cognitive load accurately. There are several ways to assess the validity of a measurement tool, including criterion-related validity, construct validity, and concurrent validity (Warner, 2013). Warner (2013) further stated that criterion-related validity refers to the degree to which the scores on a measurement tool are related to an external criterion or standard. To assess the instrument's validity, Leppink et al. compared the results of measurements from the new instrument with previously accepted means of measuring different types of cognitive load. The authors experimented with psychology students and added four questions to the instrument from previously one-measurement accepted scales. One question was Paas's scale for measuring cognitive load, another from Ayres's scale for intrinsic load, another from Czerniak's scale for extraneous load, and the last one from Czerniak's scale for the germane load (Leppink et al., 2013). The results did not reach an R2 for Czerniak's scales, so a third study was performed to test cross-validity. The third study administered the 10 questions of the new instrument to 136 psychology students and compared the results with the answers to the 10 questions of the previous group of students in the second study. Results showed that measurements for the three types of cognitive load were significantly correlated between the two studies (Leppink et al., 2013).

Construct validity refers to the degree to which a measurement tool relates to other related constructs or variables (Warner, 2013). The Leppink MES has been compared to other measures of cognitive functioning, such as memory or problem-solving ability, to assess its construct validity. Leppink et al. (2013) experimented with 56 Ph.D. students taking a statistics class to test the internal reliability and performed principal component analysis. The results showed Cronbach's alpha values of .81 for the three questions measuring intrinsic load, .75 for extraneous load, and .82 for the germane load.

Data Analysis Plan

For this quantitative quasi-experimental with a nonequivalent control group study, I conducted three separate two-tailed *t* tests, one for each type of cognitive load. I used Statistical Package for Social Sciences Version 28.0 for statistical analysis. Before I started data analysis, I cleaned the data for the two groups being compared and then checked for normality and equal variances between the two groups using visualizations and statistical tests. I then conducted the *t* tests and calculated the *t* and *p* values for each type of cognitive load. The results helped me answer the research questions. Since the cognitive load experienced by one group could be greater than or less than the cognitive load of the other group, it was appropriate to use a two-tailed *t* test. The null hypothesis for a two-tailed *t* test is that there is no difference between the means of the two groups, and the alternative hypothesis is that there is a difference (Frankfort-Nachmias & Leon-Guerrero, 2018). The *t* test calculates the *t* value, which measures the difference between the means of the two groups in terms of standard error units, and the *p* value, which is the probability of getting a *t* value as large or more significant than the one observed, assuming the null hypothesis is true (Frankfort-Nachmias & Leon-Guerrero, 2018).

Before performing the three *t* tests, I added the scores for Questions 1–3 for each survey to obtain a score for the intrinsic cognitive load experienced by the student. I then added the scores for Questions 4–6 to obtain a score for the extraneous cognitive load experienced and added Questions 7–10 to obtain the germane cognitive load score. For each *t* test, the independent variable was the method of science instruction (i.e., VRLE or traditional), and the independent variable was the score for each type of load. Since students were in a class with a teacher who chose a VRLE or in a class with a teacher who chose a traditional method, the two groups were independent of each other. I checked the following statistical assumptions:

1. Independence: The observations in each group should be independent.
2. Normality: The data in each group should be approximately normally distributed. This assumption can be checked using visualizations such as a histogram or a normal probability plot.
3. Equal variances: The variances of the two groups should be equal. This assumption can be checked using statistical tests such as Levene's test or Bartlett's test.

4. Random sampling: The samples should be randomly selected from the population.
5. Large sample size: The sample size should be large enough to ensure that the sampling distribution of the mean is normal.

Threats to Validity

Threats to validity refer to any factors that could influence the results of a study and lead to inaccurate conclusions (Robert, 2011). Some common threats to validity in a two-tailed *t* test include selection bias, confounding variables, measurement errors, or data analysis errors (Robert, 2011). Selection bias occurs when the groups being compared are not truly representative of the population from which they were selected, and this can happen if the sample is not randomly selected or specific subgroups are underrepresented or excluded from the study (Robert, 2011). Confounding variables are related to both the independent variable and the outcome variable and can influence the relationship between the two. In the current study, a teacher's skill could be considered a cofounding variable. Measurement error occurs when the measurements used to assess the variables of interest are not accurate or unreliable, leading to inaccurate results and conclusions (Roberts, 2011). Data analysis errors occur when the data needs to be adequately cleaned, transformed, or analyzed, leading to accurate or biased results (Robert, 2011).

Threats to External Validity

External validity refers to the extent to which the results of a study can be generalized to other populations, settings, and periods (Robert, 2011). In my study, the lack of diversity in the student population at PHS limits the applicability of the results to populations of different socio-economic groups. Also, the fact that PHS is a college preparatory school limits the population to those who can maintain a 2.0 grade point average in advanced courses, which may not apply to most secondary students.

Threats to Internal Validity

Internal validity refers to the extent to which the results of a study can be attributed to the independent variable and not to other extraneous variables (Warner, 2013). In other words, it concerns the degree to which the study establishes a causal relationship between the independent and dependent variables. A limiting factor in the study is that I do not consider gender difference, which some studies suggest affects the cognitive load experienced by the student (Bevilacqua, 2017; Ibili & Billinghurst, 2019). Moreover, this study does not have a random selection of students due to using naturally occurring groups. Another area for improvement of the study is that the lessons are not only taught by different methods but also by different teachers, which could be a confounding factor. It is possible that a teacher's skill or experience also influences the variables. Moreover, since the students attend the same school, they may interact with each other, which could cause diffusion of treatment.

Ethical Procedures

Several ethical issues should be considered when using archival data in a quantitative research study. My first concern was to obtain a preliminary authorization to use the data from the PHS principal, which I obtained from the school's principal and from the Internal Review Board of Walden University. I masked the identity of the site where the data were collected. The science department did not collect participants' identities. Transparency is another concern when using archival data, so I provided the IRB with the detailed process of obtaining the data and my role as an employee of PHS. The data quality was examined according to the necessary assumptions of the analytical processes. I protected the data by storing it in my personal Microsoft One Drive, which is protected by Microsoft's encryption. Each file is encrypted with a unique AES256 key, and these unique keys are encrypted with a set of master keys stored in Azure Key Vault. Walden IRB granted full approval (05-05-23-0761070), and I was permitted to proceed with my study.

Summary

In Chapter 3, I explained the rationale for using a quantitative quasi-experimental design study with non-equivalent groups. I also provided a detailed explanation of the methodology, including a description of the research population, the sampling and sampling procedures, and procedures for using archival data. Further details were provided for the instrument, Leppink's Mental Effort Survey, which was used to collect the data with supportive studies that provided information on its validity and reliability. Next, details regarding my data analysis plan were discussed, followed by explanations of the threats to external and internal validity and how these were addressed in my study. The chapter concluded with the ethical considerations of my study and how I addressed them.

Chapter 4: Results

The purpose of this quantitative quasi-experimental study was to determine the difference in cognitive load as measured with the MES between high school students who used a VRLE during science instruction for one lesson and students who did not use a VRLE. To accomplish this, I conducted a comprehensive analysis by comparing the means of the two groups for each specific type of cognitive load. By scrutinizing the data and employing statistical techniques, I sought to discern any potential increase or decrease in cognitive load experiences between the two groups. The research questions and hypotheses that guided this study were:

RQ1: What is the difference in intrinsic load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE?

$H_0$1: There is no difference in intrinsic load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

$H_1$1: There is a statistical difference in intrinsic load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

RQ 2: What is the difference in extraneous load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE?

$H_0 2$: There is no difference in extraneous load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

$H_2 2$: There is a statistical difference in extraneous load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

RQ 3: What is the difference in germane load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE?

$H_0 3$: There is no difference in germane load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

$H_3 3$: There is a statistical difference in germane load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

In Chapter 4, I present the outcomes of this study. The chapter commences with a detailed account of the data collection process, encompassing relevant information about the student demographics, which I obtained from the archival data, and providing insights into the demographics of the student sample under examination. Subsequently, in this chapter, I present the findings through the use of descriptive statistics, thoroughly addressing the assumptions and data associated with each of the three pairs of research hypotheses. I end the chapter with a comprehensive summary of the obtained findings, offering a concise overview of the results and their implications.

Data Collection

Due to the utilization of archival data in this study, I did not actively recruit the participants because the nature of the type of study relies on preexisting data sources. Instead, I obtained Walden University IRB approval to access the archival data from the MES data collected from November 9, 2021, to November 26, 2021, by the science department of the study site. I obtained the MES scores for each type of cognitive load from the traditional instruction group (n = 141) and the VRLE group (n = 82), for a total group of 223 students. Due to the archival data used in this study, there were no noted discrepancies in the data collection process. When conducting the initial data analysis on the entire data set (N = 223), I explored descriptive statistics, revealing outliers in the extraneous load variable for both groups (see Figure 1) and in the germane load variable (see Figure 2). Consequently, I removed the outliers and excluded them from each group to meet the necessary assumptions. The analysis proceeded with the remaining traditional group (n = 134) and the VRLE group (n = 73).

Figure 1

Boxplots Showing Outliers for the Extraneous Load Variable

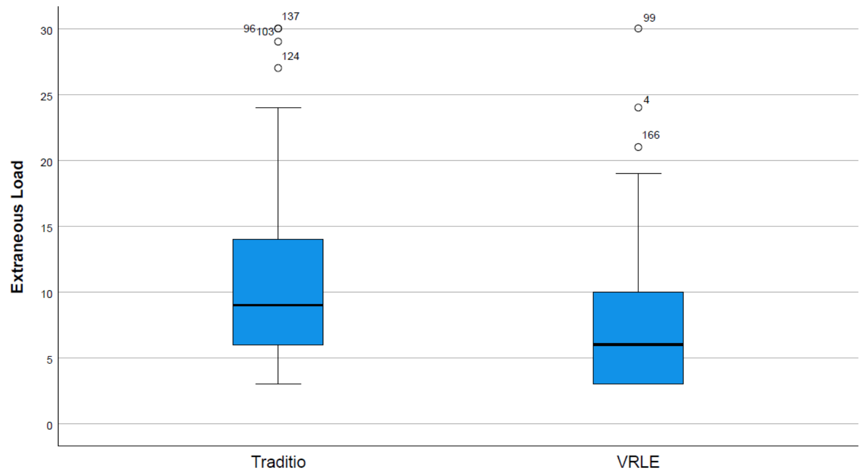
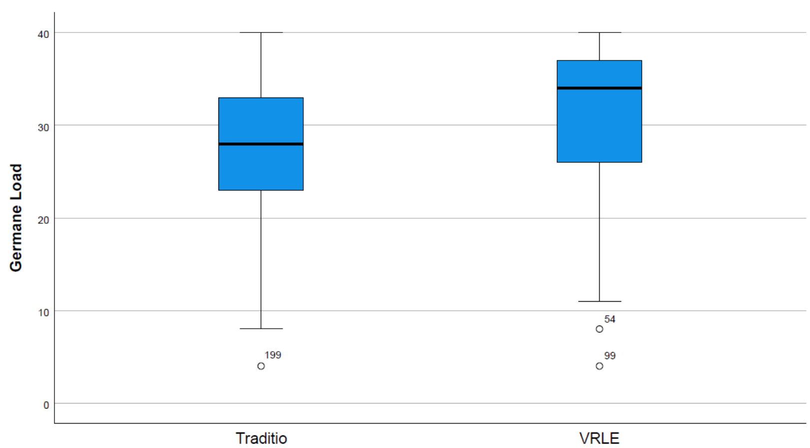


Figure 2

Boxplots Showing Outliers for the Germane Load Variable



I obtained the archival data from PHS, a private, college preparatory high school in a high socioeconomic area. The external validity of the results is limited to these types of schools. The data did not include demographics beyond the age and gender of student, with 47% of the sample female and 52% male and with ages ranging from 14–18 years old, with 43% of the sample being 15 years old. The population at the study site represents the surrounding area, and the students that took the survey were large enough to represent PHS well. There was a good representation of both genders and all high school ages, which makes the results applicable to a wide range of students in similar schools.

Results

Descriptive statistics that characterize the sample include, as shown in Table 2, the mean score for each type of cognitive load for both groups, the traditional (n = 132) and the VRLE (n = 73), and the standard deviation. Data consisted of the MES scores for each type of cognitive load for both groups. The traditional group in the study had an average intrinsic load score of 18.85, while the VRLE group had an average intrinsic load score of 15.93. Regarding extraneous load, the traditional group had an average score of 10.10, whereas the VRLE group had an average score of 6.84. Finally, the traditional group had an average germane load score of 27.98, while the VRLE group had an average germane load score of 32.68. Overall, the intrinsic and extraneous load scores were lower for the VRLE group, and the germane load scores were higher.

Table 2

Descriptive Statistics Cognitive Load

	Group	*M*	*SD*	Min.	Max.
Intrinsic load	Traditional	18.85	5.503	3	30
	VRLE	15.93	5.29	4	27
Extraneous load	Traditional	10.10	5.527	3	24
	VRLE	6.84	3.877	3	17
Germane load	Traditional	27.98	7.086	8	40
	VRLE	32.68	6.453	17	40

According to Laerd Analytics (2023), there are six assumptions for an independent samples *t* test. The first assumption is that a continuous dependent variable be used, which I met by using scores from the MES, as described in Chapter 3. The second assumption is that there is a categorical independent variable, which is the case for the method of instruction variable, which was either VRLE or traditional. The third assumption is the independence of observation, which the data in the current study also met since students were in different classes when they completed the survey. The fourth assumption is that there should be no significant outliers, so I removed the abovementioned outliers.

Using Statistical Package for Social Sciences 28.0, I tested for the assumption of normality. Table 3 displays the normal distribution of intrinsic load scores in the traditional group (Shapiro-Wilk test $p = .086$) and the VRLE group (Shapiro-Wilk test $p = .659$). However, neither the extraneous nor the germane load scores exhibited a normal distribution for either group (Shapiro-Wilk test $p < 0.001$).

Table 3

Shapiro-Wilk Test Results

	Group	Statistic	*df*	Sig.
Intrinsic load	Traditional	0.983	134	0.086
	VRLE	0.987	73	0.659
Extraneous load	Traditional	0.932	134	<.001
	VRLE	0.867	73	<.001
Germane load	Traditional	0.978	134	0.033
	VRLE	0.897	73	<.001

According to Laerd Statistics (2023), it is important to note that with larger sample sizes, thanks to the central limit theorem, the independent-sample *t* test can still yield valid results despite deviations from normality. Nonetheless, the Q-Q plots show that the data for the extraneous load does not deviate significantly from the normal line (see Figure 3) or does the data for the germane load (see Figure 4).

Figure 3

Q-Q Plots for Extraneous Load Variable

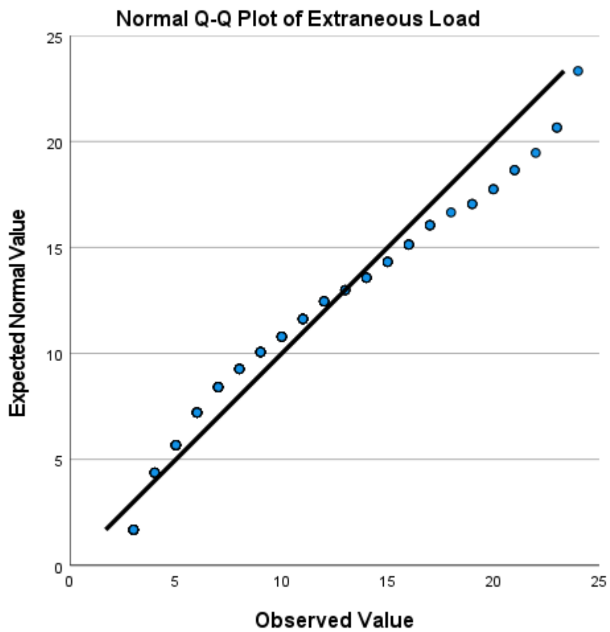
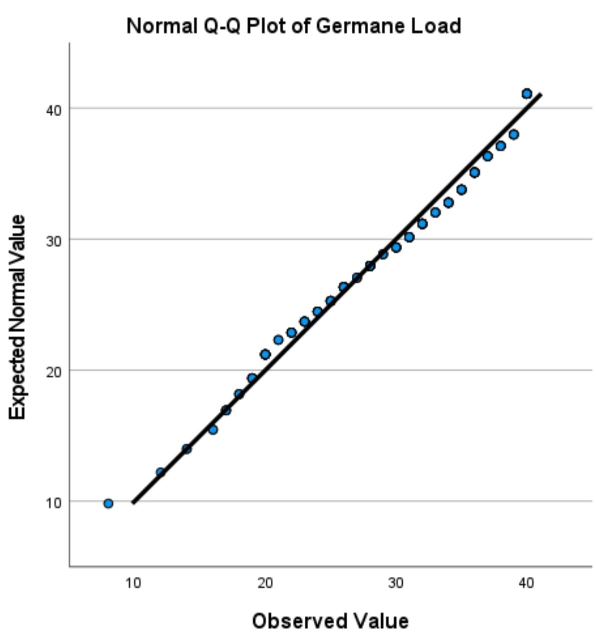


Figure 4

Q-Q Plots for Germane Load Variable



The final assumption for the independent *t* test is the homogeneity of variances, which I checked using Levene's test for equality of variances, as shown in Table 4. The intrinsic load score met the assumption (p = .598) as did the germane load score (p = .503), but the extraneous load score (p > .05) did not meet the assumption. However, I ran the *t* tests for data that both met the assumption and for data that did not meet it.

Table 4

Levene's Test for Equality of Variances

		F	Sig.
Intrinsic load	Equal variances assumed	0.278	0.598
Extraneous load	Equal variances assumed	12.337	0.001
Germane load	Equal variances assumed	0.451	0.503

Intrinsic Load

I ran an independent *t* test to discern if there were any differences in MES scores for the intrinsic load between students who used a VRLE during science instruction for one semester and students who did not use a VRLE. The mean intrinsic load score for the VRLE group was 15.93, while the mean for the non-VRLE group (i.e., traditional) was 18.85. The *t* test showed that this difference is significant with p < .001 (see Table 5), so the null hypothesis could be rejected. The first hypothesis can be answered to say not only that there is a difference but that there is a decrease in the intrinsic load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

Extraneous Load

I ran an independent *t* test to discern if there were any differences in MES scores for the extraneous load between students who used a VRLE during science instruction for one semester and students who did not use a VRLE. The mean extraneous load score for the VRLE group was 6.84, while the mean for the non-VRLE group (i.e., traditional) was 10.10. The *t* test showed that this difference is significant with $p < .001$ (see Table 5), so the null hypothesis could be rejected. The second hypothesis can be answered by saying that there is a difference in MES scores and a decrease in the extraneous load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

Germane Load

I ran an independent *t* test to discern if there were any differences in MES scores for the germane load between students who used a VRLE during science instruction for one semester and students who did not use a VRLE. The mean germane load score for the VRLE was 32.68, while the mean for the non-VRLE group (i.e., traditional) was 27.98. The *t* test showed that this difference is significant with $p < .001$ (see Table 5), so the null hypothesis could be rejected. The third hypothesis can be answered by saying there is a difference and an increase in the germane load MES scores between high school students who used a VRLE during science instruction for one semester and students who did not use a VRLE.

Table 5

t *-test for Equality of Means for the MES Scores*

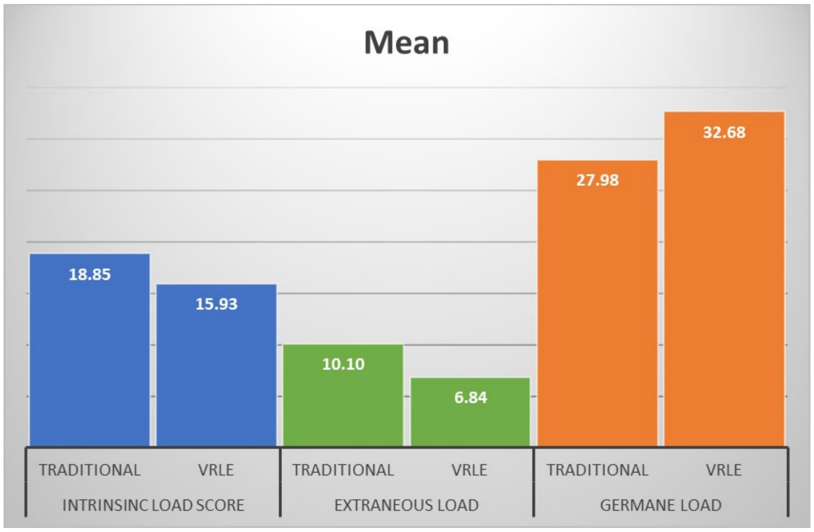
		t	df	Significance	
				One-sided p	Two-sided p
Intrinsic load	Equal variances assumed	3.697	205	< .001	< .001
	Equal variances not assumed	3.74	153.105	< .001	< .001
Extraneous load	Equal variances assumed	4.486	205	< .001	< .001
	Equal variances not assumed	4.963	192.173	< .001	< .001
Germane load	Equal variances assumed	-4.71	205	< .001	< .001
	Equal variances not assumed	-4.842	160.249	< .001	< .001

Summary

In Chapter 4, I presented this study's key findings and the assumptions associated with an independent samples *t* test. I designed the study to meet the first three assumptions. After removing the outliers, the box plot chart showed the data met the fourth assumption. A descriptive exploration of the remaining data from this study showed that all the variables did not meet the normality assumption according to the Shapiro-Wilk results; however, the Q-Q showed that it approached normality. Levene's test for equality of variances showed that the data did not meet the assumption for the homogeneity of variances, but *t* tests were run for both possibilities. Overall, this study's key findings and data answered the three RQs by indicating significant differences between the MES scores for students taught with VRLE and those not (see Figure 5).

Figure 5

Comparison of Means Between the Groups



I will begin Chapter 5 by restating the purpose and nature of the study. I will provide an overview of the methodology employed and explain the rationale behind conducting this research. Furthermore, the key findings from Chapter 4 will be summarized and connections between these findings and the current state of literature in the discipline will be established. Next, I will delve into a comprehensive discussion of the limitations of this study before making recommendations for future research in the field of science education and VR. Chapter 5 will conclude with a description of the potential social impact this study may result in and the contributions this work can make to the field.

Chapter 5: Discussion, Conclusions, and Recommendations

The problem addressed through this study was that it is not yet fully understood what the effect of VRLEs is on students' cognitive load, which can hinder their learning if it is too high. The purpose of this quantitative quasi-experimental study was to determine the difference in cognitive load as measured with the MES between high school students who used a VRLE during science instruction for one lesson and students who did not use a VRLE. I used the results from the MES, which has 10 questions: Questions 1, 2, and 3 measure intrinsic cognitive load; Questions 4, 5, and 6 measure the extraneous load; and Questions 7, 8, 9, and 10 measure germane load. I used an independent *t* test to determine if the VRLE group's mean is significantly greater than or less than the traditionally taught group. The study population was adolescents taking a secondary science course and completing the MES as part of a science lesson. The results showed that there was a decrease in intrinsic and extraneous loads between the VRLE group and the traditional group. The findings also showed an increase in germane load in the VRLE group. These results help explain why previous studies that focused only on a single total cognitive load score had conflicting results.

Interpretation of the Findings

I interpreted the cognitive effects studied using the CLT. CLT has been used to measure the effectiveness of an instructional tool in educational technology since its inception in 1988 (Sweller, 2020). In CLT, an efficient instructional tool is defined as lessening the intrinsic load, eliminating extraneous load, and focusing all mental effort on the germane load so deep learning can occur (Sweller, 2020). Therefore, the cognitive load must be measured separately to determine its effectiveness correctly. The findings of this study add clarity to the literature by confirming the studies that found VRLEs as an effective tool for increasing comprehension of complex subjects. However, although the extraneous load is also decreased in the VRLE group, it was not eliminated, leaving room for improvement, as concluded by other studies. The increase in germane load shows that students had more room for deep thought, which aligns with those studies that showed improved learning scores. I organized my discussion of the results by each type of cognitive load and the key finding for each.

Intrinsic Load

Intrinsic load arises from the complexity of the subject learned and the lack of previous knowledge held by the learner (Leppink et al., 2013). A review of the literature showed that science education has high levels of inherent complexity (Achor et al., 2019; Aguilera & Perales-Palacios, 2020; Etobro & Fabinu, 2017; Hidayati et al., 2020; Santi & Gorghiu, 2019), which caused students to lose interest in the subject matter. However, the literature also has shown that VRLEs increase students' interest in science (Astuti et al., 2020; Garduño et al., 2021), which can be explained by the decrease in intrinsic load. The findings of the current study extend the literature by explaining the relation between the use of VRLEs and the increase in students' interest in science that the current literature shows (see Aguilera & Perales-Palacios, 2020; Frazier et al., 2021; W. H. Lee et al., 2021; Makransky, Mayer, et al., 2019; Makransky, Terkildsen, et al., 2019; Parong & Mayer, 2018, 2021; Petersen et al., 2020). This study provides further evidence of the ability of VRLEs to make intangible and inaccessible concepts tangible (see Nersesian et al., 2019a; Petkov et al., 2019), such as those covered in secondary science education (Achor et al., 2019; Aguilera & Perales-Palacios, 2020; Etobro & Fabinu, 2017; Hidayati et al., 2020; Santi & Gorghiu, 2019) and that VRLEs lower students' intrinsic load. Sweller (2016) divided human knowledge into primary knowledge that is learned naturally without the need of a teacher and only causing natural levels of germane load and secondary knowledge that needs to be taught, and if not taught properly it can cause cognitive overload. Dan and Reiner (2017) explained that learning in a 3D environment mimics how humans learn naturally, tapping into the biological tools of the human person to learn intuitively, which explains why it decreases intrinsic load.

Extraneous Load

Extraneous load stems from instruction elements that do not help the student process information (Sweller, 1988). These extraneous elements can arise from the instructional design and external sources (Paas & Van Merriënboer, 1994). The current study results show that a VRLE does cause some extraneous load, which could be caused by the amount of nonessential details required to create a VRLE (see Frazier et al., 2021; Makransky, Terkildsen, et al., 2019; Zhao et al., 2020); however, the non-VR group experienced higher extraneous cognitive load. These findings may shed light on the discrepancy in results of studies conducted with K–12 populations compared to those done with university students (see Bogusevschi et al., 2018; Jitmahantakul & Chenrai, 2019; Lai et al., 2022; Liou & Chang, 2018; Meyer et al., 2019; Nersesian et al., 2019b; Parong & Mayer, 2018; Southgate, 2019). Studies that measured cognitive load in adult populations reported that the non-VRLE groups experienced less overall cognitive load and scored better in learning outcomes (Makransky, Terkildsen, et al., 2019; Parong & Mayer, 2018), while the studies with younger populations showed the opposite (Bogusevschi et al., 2018; Jitmahantakul & Chenrai, 2019; Lai et al., 2022; Liou & Chang, 2018; Nersesian et al., 2019b; Southgate, 2019). The difference between a high school classroom and a university class could explain this increase in the non-VRLE group in the current study findings. The distractions of the high school environment may pose a more significant extraneous load than those in the internal design of the VRLE. It is also possible that the traditional instructional methods used by the non-VRLE teachers caused the increase, but it is less likely given that nine different teachers gave the survey.

Germane Load

The third type of cognitive load is the germane load, which refers to the effort required to transfer new information from the working to long-term memory and properly integrate it with the existing knowledge (Leppink et al., 2013). The germane load is the desired type of cognitive load, and if an increased germane load causes an overall increase in cognitive load, then the educational tool would have the desired effect. The increased germane load by the VRLE group in the current study confirms the results of Vesga et al. (2021) who showed that the increased sense of presence created by VRLEs led students to exert more effort in comprehending the lesson. Moreover, the current study findings show a decrease in intrinsic and extraneous load, which leaves more room for the germane load.

An instructional tool can also aid in increasing the germane load by integrating and organizing new content so that the learner can store the information in long-term memory more easily (Leppink et al., 2013). Several of the studies in the literature focused on how VRLEs provide an environment where students can explore and interact with the content that is not limited by a lack of supplies or expensive equipment, features that allow students to develop a more excellent scientific attitude (Liou & Chang, 2018; Makransky & Lilleholt, 2018; Nersesian et al., 2019b; Parong & Mayer, 2021), and this is more than just increased interest, it is an attitude of the mind open to inquiry and problem solving (Astuti et al., 2020). The current study confirms that VRLEs help students create the scientific reasoning required to comprehend the subject matter, thus increasing germane load.

In the literature review, I showed that science education is inherently complex, and students are losing interest in secondary science classes (Achor et al., 2019; Aguilera & Perales-Palacios, 2020; Bal-Incebacak et al., 2019; Etobro & Fabinu, 2017; Hidayati et al., 2020; Lodge, 2021; Nidup et al., 2021; Santi & Gorghiu, 2019; Uçar & Sungur, 2017). The adolescent mind often comes into the science classroom lacking the mental schemas necessary for integrating complex information with prior existing knowledge (Bal-Incebacak et al., 2019; Sadler et al., 2013; Santi & Gorghiu, 2019; Wei et al., 2021; L. Zhang, 2019). The current study results show that a VRLE is an efficient tool that can help secondary students' cognitive processes in high school science classrooms.

Limitations of the Study

I designed this study with CLT as its theoretical foundation. CLT has a narrow perspective on instructional design, mainly emphasizing the reduction of intrinsic and extraneous cognitive load without fully considering other crucial aspects like learner motivation, interest, and engagement (Kirschner, 2011). Therefore, there could be other elements that should be considered before adopting a VRLE as an instructional tool. Moreover, CLT does not account for individual differences among learners, such as their abilities, prior knowledge, learning styles, or motivation, which can also impact the effectiveness of an instructional design.

Limitations of this study also arose from the lack of attributes of the population provided by the archival data set. The lack of ethnic or socioeconomic differences makes the findings less applicable to real-life educational contexts. The design did not consider gender differences, and much previous research has focused on how gender affects cognitive processes (Bevilacqua, 2017). This study was limited by not measuring confounding variables such as prior knowledge, a particular course or lesson, or the teacher's effectiveness.

Recommendations

My first recommendations are based on study results. The results showed an increase in the intrinsic load of the non-VRLE group, and this difference may be due to a difference in the student's prior knowledge. Therefore, more research must be done to control the prior knowledge variable. Results also included an increase in the extraneous load of the non-VRLE group, which a lack of teaching experience in classroom management or the ineffective traditional material could have caused. Therefore, further studies that focus on the same teacher could help control these variables.

Moreover, in this study I showed that some of the cognitive effects of CLT applied to VRLE, while others did not, and the results showed that the extraneous load was still present in the VRLE group, so further studies should focus on the design principles of VRLE and research methods of lessening the extraneous load it causes.

I conducted this study with a total population of 207 students between the ages of 14–18 attending a college preparatory high school in a high socioeconomic area. Therefore, this study could be replicated by other secondary schools in a similar area using the same MES to determine if the results are similar. In addition, this study could be improved by using MES scores combined with such a measure of learning outcomes, especially a design that measures immediate recall and the long-term retention of the information learned.

Implications

This study may contribute to positive social change in several ways. First, at the individual level, as more teachers make research-based and informed decisions about instruction, students' academic success may increase. There is also potential for change at the organizational level because secondary school administrators can use data from this study to make informed decisions regarding purchasing the hardware and software for science instruction. This study also may advance knowledge in educational technology by providing further evidence to the question of the effectiveness of VRLEs as a mode of science instruction for secondary students. At a societal level, the increased interest in science brought about by VRLEs may increase the number of science and technology skilled workers available as more high school students choose scientific careers for their higher education.

Another contribution that this study makes to positive social change is through increased access to advanced science courses and field trips. The results of this study indicate that VRLEs have a positive cognitive effect on students, so school districts and organizations can invest in the technology to help schools in rural areas and lower socioeconomic districts where building science labs is not a viable option.

Conclusion

VRLEs are becoming a popular innovative learning mode for K–12 institutions, and there is a growing amount of literature on the subject. Previous studies regarding VRLEs in the science classroom produced conflicting results on its effect on student performance and cognitive load (see Bogusevschi et al., 2018; Jitmahantakul & Chenrai, 2019; Lai et al., 2022; Liou & Chang, 2018; Meyer et al., 2019; Parong & Mayer, 2018; Nersesian et al., 2019b; Southgate, 2019); however, there was consensus regarding VRLEs' ability to create a sense of presence that increased student interest in science (Parong & Mayer, 2021). The question remained whether this increase in interest in science leads only to emotional engagement or also to cognitive engagement (Parong & Mayer, 2021). Results from the current study showed that this sense of presence increases the students' cognitive engagement with science, as demonstrated by decreases in intrinsic and extraneous cognitive load and an increase in the germane load. It is important for schools to invest in technology that has proven results, and this study shows that purchasing VRLE technology for high-school science courses aids in the cognitive processes of the students.

Educational technology is an ever-changing field as countless new products appear in the marketplace on an almost daily basis and schools spend billions on software and tools that promise to help students (EdTech Evidence Exchange, 2021). However, the speed at which new items appear in this market sometimes prevents a thorough vetting before they are implemented, and money is spent without seeing any benefits or sometimes even causing harm (Boston, 2021). This study provided evidence on the cognitive benefits of VRLEs but also showed that there is room for improvement as designers aim to reduce the extraneous load that the systems produce.

References

Achor, E. E., Danjuma, I. M., & Orji, A. B. C. (2019). Classroom interaction practices and students' learning outcomes in physics: Implication for teaching-skill development for physics teachers. *Journal of Education and E-Learning Research*, *6*(3), 96–106. https://doi.org/10.20448/journal.509.2019.63.96.106

Aguilera, D., & Perales-Palacios, F. J. (2020). Learning biology and geology through a participative teaching approach: The effect on student attitudes towards science and academic performance. *Journal of Biological Education*, *54*(3), 245–261. https://doi.org/10.1080/00219266.2019.1569084

Akman, E., & Çakır, R. (2019). Pupils' opinions on an educational virtual reality game in terms of flow experience. *International Journal of Emerging Technologies in Learning*, *14*(15), 121. https://doi.org/10.3991/ijet.v14i15.10576

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

Ariali, S., & Zinn, B. (2021). Adaptive training of the mental rotation ability in an immersive virtual environment. *International Journal of Emerging Technologies in Learning*, *16*(9), 20–39. https://doi.org/10.3991/ijet.v16i09.18971

Astuti, T. N., Sugiyarto, K. H., & Ikhsan, J. (2020). Effect of 3D visualization on students' critical thinking skills and scientific attitude in chemistry. *International Journal of Instruction*, *13*(1), 151–164. http://10.0.114.149/iji.2020.13110a

Aubteen Darabi, A., Nelson, D. W., & Paas, F. (2007). Learner involvement in instruction on a complex cognitive task: Application of a composite measure of performance and mental effort. *Journal of Research on Technology in Education*, *40*(1), 39–48. https://doi.org/10.1080/15391523.2007.10782495

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

Baceviciute, S., Terkildsen, T., & Makransky, G. (2021). Remediating learning from non-immersive to immersive media: Using EEG to investigate the effects of environmental embeddedness on reading in virtual reality. *Computers and Education*, *164*, 104122. https://doi.org/10.1016/j.compedu.2020.104122

Bal-Incebacak, B., Yaman, S., & Sarisan-Tungaç, A. (2019). The relation between intellectual risk-taking regarding science classes and test anxiety inventory of secondary school. *South African Journal of Education*, *39*(1).

Becker, S., Klein, P., Gößling, A., & Kuhn, J. (2020). Using mobile devices to enhance inquiry-based learning processes. *Learning and Instruction*, *69*. https://doi.org/10.1016/j.learninstruc.2020.101350

Bevilacqua, A. (2017). Commentary: Should gender differences be included in the evolutionary upgrade to cognitive load theory? *Educational Psychology Review*, *29*(1), 189–194. https://doi.org/10.0.3.239/s10648-016-9362-6

Boda, P. A., & Brown, B. (2020). Designing for relationality in virtual reality: Context-specific learning as a primer for content relevancy. *Journal of Science Education and Technology*, *29*(5), 691–702. https://doi.org/10.1007/s10956-020-09849-1

Bogusevschi, D., Muntean, C. H., & Muntean, G.-M. (2020). Teaching and learning physics using 3D virtual learning environment: A case study of combined virtual reality and virtual laboratory in secondary school. *Journal of Computers in Mathematics & Science Teaching*, *39*(1), 5. https://www.learntechlib.org/primary/p/210965/

Bogusevschi, D., Tal, I., Bratu, M., Gornea, B., Caraman, D., Ghergulescu, I., Hava Muntean, C., & Muntean, G.-M. (2018). Water cycle in nature: Small-scale STEM education pilot. *EdMedia + Innovate Learning, 2018*, 1496–1505. https://www.learntechlib.org/p/184370

Boston, W. (2021). *Are expenditures for EdTech in K-12 education wasteful?* https://wallyboston.com/are-expenditures-for-edtech-in-k-12-education-wasteful/

Burkholder, G., Cox, K., & Crawford, L. (2016). *The scholar-practitioner's guide to research design* (Kindle ed.). Laureate Publishing.

Cai, S., Chiang, F. K., Sun, Y., Lin, C., & Lee, J. J. (2017). Applications of augmented reality-based natural interactive learning in magnetic field instruction. *Interactive Learning Environments*, *25*(6), 778–791. https://doi.org/10.1080/10494820.2016.1181094

Cates, W. M. (1992). Considerations in evaluating metacognition in interactive hypermedia/multimedia instruction (ED349966). *Annual Meeting of the American Educational Research Association*, ERIC. 1–17. https://eric.ed.gov/?id=ED349966

Chen, C. J. (2006). The design, development and evaluation of a virtual reality based learning environment. *Australasian Journal of Educational Technology*, *22*(1). https://doi.org/10.14742/ajet.1306

Chen, O., Manalo, E., & She, Y. (2019). Examining the influence of expertise on the effectiveness of diagramming and summarizing when studying scientific materials. *Educational Studies*, *45*(1), 57–71. https://doi.org/10.1080/03055698.2017.1390444

Chen, R. (2016). Learner perspectives of online problem-based learning and applications from cognitive load theory. *Psychology Learning and Teaching*, *15*(2), 195–203. https://doi.org/ 10.1177/1475725716645961

Cheng, K. H., & Tsai, C. C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers and Education*, *140*(June), 103600. https://doi.org/10.1016/j.compedu.2019.103600

Chilton, M., & Gurung, A. (2008). Management of lecture time: Using the web to manipulate extrinsic cognitive load. *International Journal of Web-Based Learning and Teaching Technologies*, *3*(2), 35. https://doi.org/10.4018/jwltt.2008040103

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction. Proven guidelines for consumers and designers of multimedia learning* (4th ed.). Pfeiffer.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* Academic Press.

Coleman, R. (2022). *Designing experiments for the social sciences: How to plan, create, and execute research using experiments.* Sage Publications. https://doi.org/10.4135/9781071878958

Cook, D. A., Castillo, R. M., Gas, B., & Artino, A. R. (2017). Measuring achievement goal motivation, mindsets and cognitive load: Validation of three instruments' scores. *Medical Education*, *51*(10), 1061–1074. https://doi.org/10.1111/medu.13405

Cosgrove, S. B., & Olitsky, N. H. (2018). From "traditional" to research-based instructional strategies: An assessment of learning gains. *AEA Papers and Proceedings*, *108*, 302–306. https://doi.org/10.1257/PANDP.20181053

Dan, A., & Reiner, M. (2017). EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. *International Journal of Psychophysiology*, *122*, 75–84. https://doi.org/10.1016/j.ijpsycho.2016.08.013

Daniel, J. (2012). *Sampling essentials: Practical guidelines for making sampling choices.* https://doi.org/10.4135/9781452272047

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, *100*(1), 223–234. https://doi.org/10.1037/0022-0663.100.1.223

Dong, A., Jong, M. S. Y., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/FPSYG.2020.591203/FULL

EdTech Evidence Exchange. (2021). *Overview: U.S. K-12 public education technology spending*. https://edtechevidence.org/wp-content/uploads/2021/07/FINAL-K12-

EdTech-Funding-Analysis_v.1.pdf

Etobro, A. B., & Fabinu, O. E. (2017). Students' perceptions of difficult concepts in biology in senior secondary schools in Lagos state. *Global Journal of Educational Research*, *16*(2), 139. https://doi.org/10.4314/gjedr.v16i2.8

Frankfort-Nachmias, C., & Leon-Guerrero, A. (2018). *Social statistics for a diverse society* (8th ed.). SAGE Publications.

Frazier, E., Lege, R., & Bonner, E. (2021). Making virtual reality accessible for language learning: Applying the VR application analysis framework. *Teaching English with Technology*, *21*(1), 128–140.

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., & Andersen, S. A. W. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: A randomized trial. *Surgical Endoscopy*, *34*(3), 1244–1252. https://doi.org/10.1007/s00464-019-06887-8

Garduño, H. A. S., Martínez, M. I. E., & Castro, M. P. (2021). Impact of virtual reality on student motivation in a high school science course. *Applied Sciences*, *11*(20). https://doi.org/10.3390/app11209516

Gielstra, D., Moorman, L., Cerney, D., Cerveny, N., & Gielstra, J. (2021). Geoepic: Innovating a solution to implement virtual field experiences for education in the time of COVID-19 and the post-pandemic era. *Journal of Higher Education Theory and Practice*, *21*(7), 1–10. https://doi.org/10.33423/JHETP.V21I7.4481

Hadie, S. N. H., Hassan, A., Mohd Ismail, Z. I., Ismail, H. N., Talip, S. B., & Abdul

Rahim, A. F. (2016). Empowering students' minds through a cognitive load theory-based lecture model: A metacognitive approach. *Innovations in Education and Teaching International*, 1–10. https://doi.org/10.1080/14703297.2016.1252685

Hatchard, T., Azmat, F., Al-Amin, M., Rihawi, Z., Ahmed, B., & Alsebae, A. (2019). Examining student response to virtual reality in education and training. *2019 IEEE 17th International Conference on Industrial Informatics*, *1*, 1145–1149. https://doi.org/10.1109/INDIN41052.2019.8972023

Hidayati, N., Tanah Boleng, D., & Candra, K. P. (2020). Students' learning motivation and cognitive competencies in the PP and PBL models. *Journal Pendidikan Biologi Indonesia*, *6*(3), 367–374. https://doi.org/10.22219/jpbi.v6i3.12081

Hochberg, K., Becker, S., Louis, M., Klein, P., & Kuhn, J. (2020). Using smartphones as experimental tools—a follow-up: Cognitive effects by video analysis and reduction of cognitive load by multiple representations. *Journal of Science Education and Technology*, *29*(2), 303–317. https://doi.org/10.1007/S10956-020-09816-W

Hu Au, E., & Lee, J. J. (2017). Virtual reality in education: A tool for learning in the experience age. *International Journal of Innovation in Education*, *4*(4), 215. https://doi.org/10.1504/ijiie.2017.10012691

Huang, C. L., Luo, 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

Huang, H.-M., & Liaw, S.-S. (2018). An analysis of learners' intentions toward virtual reality learning based on constructivist and technology acceptance approaches. *The International Review of Research in Open and Distributed Learning*, *19*(1), 91–115. https://doi.org/10.19173/irrodl.v19i1.2503

Huynh, T., Hou, G., & Wang, J. (2016). Communicating wave energy: An active learning experience for students. *American Journal of Engineering Education*, *7*(1), 37–46. https://doi.org/10.19030/ajee.v7i1.9684

Ibili, E., & Billinghurst, M. (2019). Assessing the relationship between cognitive load and the usability of a mobile augmented reality tutorial system: A study of gender effects. *International Journal of Assessment Tools in Education*, *6*(3), 378–395. https://doi.org/10.21449/ijate.594749

Jitmahantakul, S., & Chenrai, P. (2019). Applying virtual reality technology to geoscience classrooms. *Review of International Geographical Education Online*, *9*(3), 577–590. https://doi.org/10.33403/rigeo.592771

Johnson, N. (2012). Examining self regulated learning in relation to certain selected variables. *Acta Didactica Napocensia*, *5*(3), 1–12.

Jonas, M. E. (2011). Dewey's conception of interest and its significance for teacher education. *Educational Philosophy and Theory*, *43*(2), 112–129. https://doi.org/10.1111/J.1469-5812.2009.00543.X

Keller, T., Glauser, P., Ebert, N., & Brucker-Kley, E. (2018). Virtual reality at secondary school – First results. *Proceedings of the 15th International Conference on Cognition and Exploratory Learning in the Digital Age, CELDA 2018*, *Celda*, 53–60.

Kenna, J. L., & Potter, S. (2018). Experiencing the world from inside the classroom: Using virtual field trips to enhance social studies instruction. *The Social Studies*, *109*(5), 265–275. https://doi.org/10.1080/00377996.2018.1515719

Khotimah, S. H., Krisnawati, N. M., Abusiri, & Budi, A. S. (2021). Contribution of virtual field trip and spatial intelligence toward the improvement in science learning achievement of elementary school students. *AIP Conference Proceedings*, *2320*. https://doi.org/10.1063/5.0037619

Kirschner, P. A., Ayres, P., & Chandler, P. (2011). Contemporary cognitive load theory research: The good, the bad and the ugly. *Computers in Human Behavior*, *27*(1), 99–105. https://doi.org/10.1016/J.CHB.2010.06.025

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*(NOV). https://doi.org/10.3389/fpsyg.2017.01997

Kokkonen, T., Lichtenberger, A., & Schalk, L. (2022). Concreteness fading in learning secondary school physics concepts. *Learning and Instruction*, *77*(August 2020), 101524. https://doi.org/10.1016/j.learninstruc.2021.101524

Krell, M. (2015). Evaluating an instrument to measure mental load and mental effort using item response theory. *Science Education Review Letters*, *2015*, 1–6. http://edoc.hu-berlin.de/serl

Küçük, S., Yilmaz, R. M., & Göktaş, Y. (2014). Augmented reality for learning English: Achievement, attitude and cognitive load levels of students. *Egitim ve Bilim*, *39*(176), 393–404. https://doi.org/10.15390/EB.2014.3595

Kumar, V., Gulati, S., Deka, B., & Sarma, H. (2021). Teaching and learning crystal structures through virtual reality based systems. *Advanced Engineering Informatics*, *50*. https://doi.org/10.1016/j.aei.2021.101362

Laerd Statistics. (2023). *Independent-samples t-test in SPSS statistics*. https://statistics.laerd.com/premium/spss/istt/independent-t-test-in-spss-7.php

Lai, T. L., Lin, Y. S., Chou, C. Y., & Yueh, H. P. (2022). Evaluation of an inquiry-based virtual lab for junior high school science classes. *Journal of Educational Computing Research*, *59*(8), 1579–1600. https://doi.org/10.1177/07356331211001579

Lamb, R., Antonenko, P., Etopio, E., & Seccia, A. (2018). Comparison of virtual reality and hands on activities in science education via functional near infrared spectroscopy. *Computers and Education*, *124*, 14–26. https://doi.org/10.1016/j.compedu.2018.05.014

Lardi, C., & Leopold, C. (2022). Effects of interactive teacher-generated drawings on students' understanding of plate tectonics. *Instructional Science*, *50*(2), 273–302. https://doi.org/10.1007/s11251-021-09567-0

Laurel, B., Oren, T., & Don, A. (1990). Issues in multimedia interface design: Media integration and interface agents. *Conference on Human Factors in Computing Systems - Proceedings*, 133–139. https://doi.org/10.1145/97243.97265

Lee, E. A. L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers and Education*, *79*, 49–58. https://doi.org/10.1016/j.compedu.2014.07.010

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

Lee, W. H., Kim, C., Kim, H., Kim, H. S., & Lim, C. (2021). Students' reactions to virtual geological field trip to Baengnyeong Island, South Korea. *ISPRS International Journal of Geo-Information*, *10*(12), 799. https://doi.org/10.3390/ijgi10120799

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

Lin, M.-C., Tutwiler, M. S., & Chang, C.-Y. (2011). Exploring the relationship between virtual learning environment preference, use, and learning outcomes in 10th grade earth science students. *Learning, Media & Technology*, *36*(4), 399–417. https://doi.org/10.1080/17439884.2011.629660

Liou, H. H., Yang, S. J. H., Chen, S. Y., & Tarng, W. (2017). The influences of the 2D image-based augmented reality and virtual reality on student learning. *Educational Technology and Society*, *20*(3), 110–121.

Liou, W. K., & Chang, C. Y. (2018). Virtual reality classroom applied to science education. *2018 23rd International Scientific-Professional Conference on Information Technology, IT 2018*, *2018-Janua*, 1–4. https://doi.org/10.1109/SPIT.2018.8350861

Lodge, W. (2021). ‘Complex and confusing’: The language demands of school science texts. *Research in Science and Technological Education*, *39*(4), 489–505. https://doi.org/10.1080/02635143.2020.1772740

Luong, T., Martin, N., Argelaguet, F., & Lecuyer, A. (2019). Studying the mental effort in virtual versus real environments. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*, 809–816. https://doi.org/10.1109/VR.2019.8798029

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, *66*(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., Mayer, R. E., Veitch, N., Hood, M., Christensen, K. B., & Gadegaard, H. (2019). Equivalence of using a desktop virtual reality science simulation at home and in class. *PLoS ONE*, *14*(4). https://doi.org/10.1371/journal.pone.0214944

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

Martin, A., Ginns, P., Burns, E., Kennett, R., Munro-Smith, V., Collie, R., & Pearson, J.

(2021). Assessing instructional cognitive load in the context of students' psychological challenge and threat orientations: A multi-level latent profile analysis of students and classrooms. *Frontiers in Psychology*, *12*, 2415. https://doi.org/10.3389/fpsyg.2021.656994

Martin, A. J., Ginns, P., Burns, E. C., Kennett, R., & Pearson, J. (2021). Load reduction instruction in science and students' science engagement and science achievement. *Journal of Educational Psychology, 113*(6), 1126–1142. https://doi.org/10.1037/edu0000552

May, P. (2005). Analysis of computer algebra system tutorials using cognitive load theory. *International Journal for Technology in Mathematics Education*, *11*(4), 117–138.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/s15326985ep3801_6

Mcpherson, H., Frank, G., Pearce, R., & Hoffman, E. (2021). Virtual field trips: Pivoting cross-curricular experiential learning to an online platform. *Science Teacher*, *88*(6), 45–51.

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers and Education*, *140*, 103603. https://doi.org/10.1016/j.compedu.2019.103603

Montgomery, L. D., Montgomery, R. W., & Guisado, R. (1995). Rheoencephalographic and electroencephalographic measures of cognitive workload: Analytical procedures. *Biological Psychology*, *40*(1–2), 143–159. https://doi.org/10.1016/0301-0511(95)05117-1

Morrison, B. B., Dorn, B., & Guzdial, M. (2014). Measuring cognitive load in introductory CS: Adaptation of an instrument. *ICER 2014 - Proceedings of the 10th Annual International Conference on International Computing Education Research*, 131–138. https://doi.org/10.1145/2632320.2632348

Murphy, P. K., Greene, J. A., Allen, E., Baszczewski, S., Swearingen, A., Wei, L., & Butler, A. M. (2018). Fostering high school students' conceptual understanding and argumentation performance in science through Quality Talk discussions. *Science Education*, *102*(6), 1239–1264. https://doi.org/10.1002/sce.21471

Naismith, L. M., Cheung, J. J. H., Ringsted, C., & Cavalcanti, R. B. (2015). Limitations of subjective cognitive load measures in simulation-based procedural training. *Medical Education*, *49*(8), 805–814. https://doi.org/10.1111/medu.12732

Nartey, E., & Hanson, R. (2021). The perceptions of senior high school students and teachers about organic chemistry: A Ghanaian perspective. *Science Education International*, *32*(4), 331–342. https://doi.org/10.33828/sei.v32.i4.8

National Center for Education Statistics, U.S. Department of Education, & Institute of Education Sciences. (2015). *NAEP report cards - Home*. http://www.nationsreportcard.gov/

National Research Council. (2000). *Inquiry and the national science education standards* (S. Olson & S. Loucks-Horsley, Eds.). National Academies Press.

https://doi.org/10.17226/9596

Nelson, B. C., Kim, Y., & Slack, K. (2016). Visual signaling in a high-search virtual world-based assessment: A SAVE Science design study. *Technology, Knowledge and Learning*, *21*(2), 211–224. https://doi.org/10.1007/s10758-016-9281-0

Nersesian, E., Spryszynski, A., & Lee, M. J. (2019). Integration of virtual reality in secondary STEM education. *2019 IEEE Integrated STEM Education Conference (ISEC)*, 83–90. https://doi.org/10.1109/ISECon.2019.8882070

New Media Consortium, & Consortium for School Networking. (2017). *Horizon report: K-12 edition, 2009-2017 | Educause*. https://library.educause.edu/resources/2017/12/horizon-report-k-12-edition-2009-2017

Nidup, Y., Zangmo, S., Rinzin, Y., Yuden, S., Subba, H. R., & Rai, J. (2021). The perception of class x students of Phuentsholing higher secondary school towards Chemistry. *Anatolian Journal of Education*, *6*(1), 51–66. https://doi.org/10.29333/aje.2021.614a

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, *32*(1–2), 1–8. https://doi.org/10.1023/B:TRUC.0000021806.17516.D0/METRICS

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8

Paas, F., & Van Merriënboer, J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review,* 6(4), 351–371.

Paas, F., Van Merriënboer, J., & Adam, J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*(1 Pt 2), 419–430. https://doi.org/10.2466/pms.1994.79.1.419

Park, B., & Brünken, R. (2015). The rhythm method: A new method for measuring cognitive load-an experimental dual-task study. *Applied Cognitive Psychology*, *29*(2), 232–243.

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. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, *37*(1), 226–241. https://doi.org/10.1111/jcal.12482

Patall, E. A., Hooper, S., Vasquez, A. C., Pituch, K. A., & Steingut, R. R. (2018). Science class is too hard: Perceived difficulty, disengagement, and the role of teacher autonomy support from a daily diary perspective. *Learning and Instruction*, *58*, 220–231. https://doi.org/10.1016/j.learninstruc.2018.07.004

Patall, E. A., Pituch, K. A., Steingut, R. R., Vasquez, A. C., Yates, N., & Kennedy, A. A. U. (2019). Agency and high school science students' motivation, engagement, and classroom support experiences. *Journal of Applied Developmental Psychology*, *62*, 77–92. https://doi.org/10.1016/j.appdev.2019.01.004

PerkinsCoie. (2018). Augmented and virtual reality survey report industry Insights into the future of AR / VR. *PerkinsCoie*, *March*, 1–22. https://www.perkinscoie.com/images/content/1/8/v2/187785/2018-VR-AR-Survey-Digital.pdf

Petersen, G. B., Klingenberg, S., Mayer, R. E., & Makransky, G. (2020). The virtual field trip: Investigating how to optimize immersive virtual learning in climate change education. *British Journal of Educational Technology*, *51*(6), 2098–2114. https://doi.org/10.1111/bjet.12991

Petkov, T., Mitkova, M., Surchev, S., Popov, S., Todorov, M., Sotirova, E., Sotirov, S., Bozov, H., Minkov, M., & Tankov, I. (2019). An application of virtual reality technology in education. *2019 29th Annual Conference of the European Association for Education in Electrical and Information Engineering*, 1–7. https://doi.org/10.1109/EAEEIE46886.2019.9000415

Price, Y., & de Ruiters, E. S. (2021). The virtual field trip: conditions of access/ibility and configurations of care in teaching ethnography (during COVID-19). *Anthropology Southern Africa*, *44*(3), 138–154. https://doi.org/10.1080/23323256.2021.2012491

Robert, C. P. (2011). A handbook of statistical analyses using R, second edition by Brian S. Everitt, Torsten Hothorn. *International Statistical Review*, *79*(2), 276–277. https://doi.org/10.1111/J.1751-5823.2011.00149_5.X

Sadler, T. D., Romine, W. L., Stuart, P. E., & Merle-Johnson, D. (2013). Game-based curricula in biology classes: Differential effects among varying academic levels.

Journal of Research in Science Teaching, *50*(4), 479–499. https://doi.org/10.1002/tea.21085

Santi, E. A., & Gorghiu, G. (2019). Cognitive and emotional dimensions recorded when implementing specific responsible research and innovation aspects in science lessons. *Revista Romaneasca Pentru Educatie Multidimensionala*, *11*(3), 224–234. https://doi.org/10.18662/rrem/147

Saw, K. G. (2017). Cognitive Load Theory and the Use of Worked Examples as an Instructional Strategy in Physics for Distance Learners: A Preliminary Study. *Turkish Online Journal of Distance Education*, 18 (4) , 142-159. htpps//doi.org/10.17718/tojde.340405

Seifan, M., Dada, O. D., & Berenjian, A. (2020). The effect of real and virtual construction field trips on students' perception and career aspiration. *Sustainability*, *12*(3), 1–14. https://doi.org/10.3390/su12031200

Settlage, J., & Southerland, S. A. (2019). Epistemic tools for science classrooms: The continual need to accommodate and adapt. *Science Education*, *103*(4), 1112–1119. https://doi.org/10.1002/SCE.21510

Shin, D., & Park, S. (2019). 3D learning spaces and activities fostering users' learning, acceptance, and creativity. *Journal of Computing in Higher Education*, *31*(1), 210–228. https://doi.org/10.1007/s12528-019-09205-2

Skulmowski, A., & Rey, G. D. (2017). Measuring cognitive load in embodied learning settings. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01191

Southgate, E. (2019). Virtual reality for deeper learning: An exemplar from high school science. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*, 1633–1639. https://doi.org/10.1109/VR.2019.8797841

Steier, R., & Kersting, M. (2019). Metaimagining and embodied conceptions of spacetime. *Cognition and Instruction*, *37*(2), 145–168. https://doi.org/10.1080/07370008.2019.1580711

Sun, R., Wu, Y. J., & Cai, Q. (2019). The effect of a virtual reality learning environment on learners' spatial ability. *Virtual Reality*, *23*(4), 385–398. https://doi.org/10.1007/s10055-018-0355-2

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2016). *Cognitive load theory, evolutionary educational psychology, and instructional design* (pp. 291–306). Springer. https://doi.org/10.1007/978-3-319-29986-0_12

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research & Development*, *68*(1), 1–16.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review,* 31(2), 261–292. Springer. https://doi.org/10.1007/s10648-019-09465-5

Torres, F., Tovar, L. A. N., & Egremy, M. C. (2015). Virtual interactive laboratory applied to high schools programs. *Procedia Computer Science*, *75*, 233–238. https://doi.org/10.1016/j.procs.2015.12.243

Uçar, F. M., & Sungur, S. (2017). The role of perceived classroom goal structures, self-efficacy, and engagement in student science achievement. *Research in Science and Technological Education*, *35*(2), 149–168. https://doi.org/10.1080/02635143.2017.1278684

Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, *41*(2), 167–174. https://doi.org/10.1111/J.1469-8986.2003.00148.X

Vesga, J. B., Xu, X., & He, H. (2021). The effects of cognitive load on engagement in a virtual reality learning environment. *Proceedings - 2021 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2021*, 645–652. https://doi.org/10.1109/VR50410.2021.00090

Warner, R. (2013). *Applied statistics: From bivariate through multivariate techniques* (2nd ed.). SAGE Publications.

Wei, L., Firetto, C. M., Duke, R. F., Greene, J. A., & Murphy, P. K. (2021). High school students' epistemic cognition and argumentation practices during small-group quality talk discussions in science. *Education Sciences*, *11*(10). https://doi.org/10.3390/educsci11100616

Weng, C., Otanga, S., Weng, A., & Cox, J. (2018). Effects of interactivity in E-textbooks on 7th graders science learning and cognitive load. *Computers and Education*, *120*, 172–184. https://doi.org/10.1016/j.compedu.2018.02.008

Yildirir, H. E. (2020). Secondary school students' initial and changes in cognitive structures of argument and related concepts. *International Journal of Research in Education and Science*, *6*(2), 231–249. https://doi.org/10.46328/ijres.v6i2.859

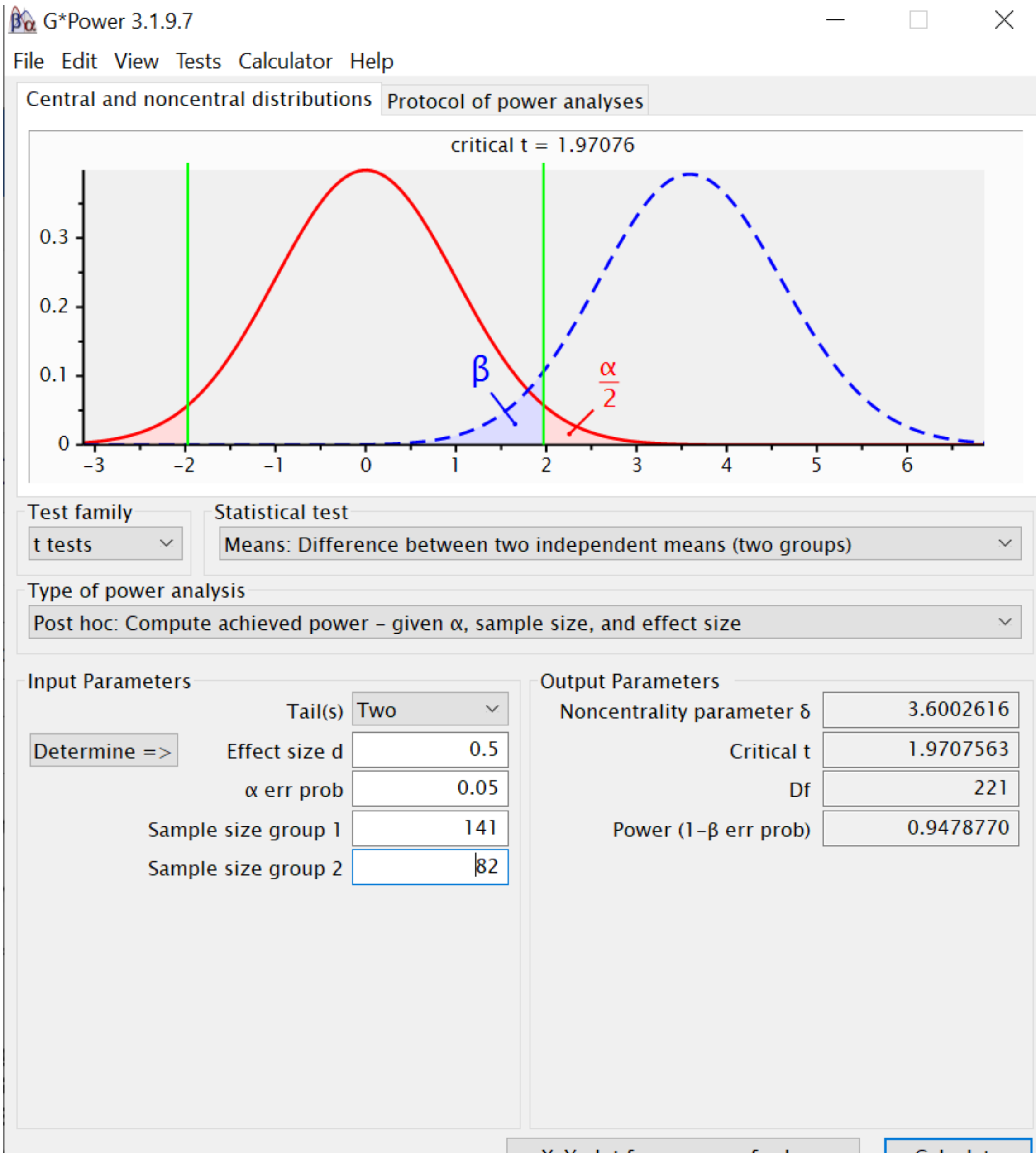
Yu, L., Holmqvist, K., Miyoshi, K., & Ashida, H. (2017). Effects of detailed illustrations on science learning: An eye-tracking study. *Instructional Science*, *45*(5), 557–581. http://10.0.3.239/s11251-017-9417-1

Zhang, H., Cui, Y., Shan, H., Qu, Z., Zhang, W., Tu, L., & Wang, Y. (2020). Hotspots and trends of virtual reality, augmented reality and mixed reality in education field. *Proceedings of 6th International Conference of the Immersive Learning Research Network, ILRN 2020*, 215–219. https://doi.org/10.23919/iLRN47897.2020.9155170

Zhang, L. (2019). "Hands-on" plus "inquiry"? Effects of withholding answers coupled with physical manipulations on students' learning of energy-related science concepts. *Learning and Instruction*, *60*, 199–205. https://doi.org/10.1016/j.learninstruc.2018.01.001

Zhao, J., Lin, L., Sun, J., & Liao, Y. (2020). Using the summarizing strategy to engage learners: empirical evidence in an immersive virtual reality environment. *Asia-Pacific Education Researcher*, *29*(5), 473–482. https://doi.org/10.1007/s40299-020-00499-w

Appendix A: Power Analysis



Power analysis of the data sample.

Appendix B: Permission to Use Survey

FW: Mental Effort Survey

Tuesday, March 23, 2021 8:30 AM

Subject	**FW: Mental Effort Survey**
Link to Outlook Item	[Click here]
From	[redacted]
To	[redacted]
Sent	**3/23/2021, 8:29:17 AM**

Sincerely,

[redacted]

Doctoral Candidate in Educational Technology Program

STUDENT ID: A00761070

From: Jimmie Leppink [redacted]
Sent: Tuesday, March 23, 2021 12:02 AM
To: Maria Carmen Lagalante [redacted]
Subject: RE: Mental Effort Survey

Dear Mrs. Lagalante,

Thank you for the message, and yes feel free to use the instrument referring to the original work.

All the best,

Jimmie

Dr Jimmie Leppink

Senior Lecturer in Medical Education, Chair of the Undergraduate Board of Examiners, and Academic Lead and Director of Assessment at Hull York Medical School, University of York