ABSTRACT

HARPER-GAMPP, TYLER. Exploring Middle School Students' Scale Cognition and Knowledge Integration During Virtual Reality-Supported Science Instruction. (Under the direction of Dr. Cesar Delgado and Dr. Karen B. Chen).

Scientific discoveries at the extremes of scale impact society across multiple levels (Winkelmann & Bhushan, 2016), establishing the need for a future workforce and citizenry able to navigate across scales. "Scale, proportion, and quantity" (SPQ) reasoning is identified by the Next Generation Science Standards (NGSS) as a crosscutting concept that pervades science (NGSS Lead States, 2013). However, research has demonstrated students' and teachers' difficulty navigating such extremes of scale (e.g., Tretter et al., 2006a, 2006b), which might be attributed to the inability to experience entities at the extremes of scale (Magaña, 2014). The theory of embodied cognition posits that there is an inextricable link between the environment, mind, and body, which suggests that experiences that foster such connections will support student learning (Wilson, 2002). However, entities (e.g., atoms) referenced throughout the NGSS cannot be experienced directly due to their immense size relative to human scale or too small to see. Therefore, virtual reality (VR) is situated to support students' size and scale reasoning by providing students with direct, embodied experiences of entities at the extremes of scale. However, to date, research exploring how and when VR supports learning is limited (Luo et al., 2021), especially at the middle school level (Matovu et al., 2023). Using an experimental explanatory mixed methods design (Creswell & Plano Clark, 2018), I explored the impact of co-developed VR-supported instruction regarding the relationship between human systems and cropland on 1) students' conceptions of size and scale, and 2) students' integration across SPQ and multiple disciplinary core ideas (DCIs). The focus on integration across a CCC and DCIs is important because the NGSS posit that the integration of CCCs with DCIs and SEPs supports students' understanding of scientific phenomena (Nordine & Lee, 2021). Participants consisted of two eighth grade classes from a Title I school located in the southeastern United States. At baseline and conclusion, students completed the Assessment of Size and Scale Cognition (ASSC) and an SPQ and DCI integration assessment. Following the intervention, a subset of students was interviewed to gain a deeper understanding of students' experience.

The enactment of the co-developed nine-day instructional unit resulted in statistically significant learning gains among students, as assessed by the ASSC. Both VR and non-VR supported instruction showed improvement, with VR-supported instruction significantly enhancing students' performance on ordering, relative, and absolute task items. A few nuanced differences emerged between the two instructional approaches, but neither emerged as the clearly better approach. Both methods notably improved students' awareness of the microscopic world. However, only the Sun and school bus were identified as collective landmarks for relative size, where at least the middle half of students estimated the order-of-magnitude correctly, among non-VR students, with VR students not achieving any collective landmarks. However, students expressed positive experiences with VR, highlighting its ability to provide embodied interactions and visualize abstract concepts.

Students integrated the CCC of SPQ with various DCIs to demonstrate the magnitude of human impacts on the environment, recognizing the long-term and large-scale impacts of local processes, in alignment with educational frameworks (NRC, 2012). Assessment items aimed at eliciting integration, however, did not necessarily result in increased integration. Instead, students tended to focus on a single scale or only referenced other scales, without describing connections across scales.

Overall, this dissertation sheds light on the potential of VR-supported instruction in fostering students' size and scale reasoning. However, further research is needed to explore how and when VR supports learning, and other techniques for capturing students' integration of CCCs with DCIs to understand how their integration supports a deeper understanding of scientific phenomena.

Exploring Middle School Students' Scale Cognition and Knowledge Integration During Virtual Reality-Supported Science Instruction

By
Tyler Harper-Gampp

A dissertation submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Learning and Teaching in STEM

Raleigh, North Carolina
2024

APPROVED BY:

Dr. Cesar Delgado Committee Co-Chair	Dr. Karen B. Chen Committee Co-Chair
Dr. M. Gail Jones	Dr. Maria Oliver-Hoyo

DEDICATION

To my loving husband, Ryan, for your unending support and guidance that pushes me to grow beyond my bounds, and for your playful inner child that always elicits joy. To my mom, Michelle, who taught me the value of education, thank you for your continued support. And to my grandma, Peggy, for your continued encouragement.

BIOGRAPHY

Tyler Harper-Gampp, born and raised in Portsmouth, Ohio, embarked on his academic journey at the University of North Carolina Wilmington. There, he earned a bachelor's degree in chemistry with a minor in mathematics and a Master of Arts in Teaching in secondary science education. His deep commitment to enhancing teaching and learning in STEM subjects ignited his aspiration to pursue a doctoral degree in STEM education. Tyler eagerly anticipates the opportunity to continue cultivating this passion as both an educator and a researcher.

ACKNOWLEDGEMENTS

With profound appreciation and admiration, I would like to express my deepest gratitude to Dr. Cesar Delgado and Dr. Karen B. Chen, my dissertation co-chairs, and to my committee members, Dr. M. Gail Jones and Dr. Maria Oliver-Hoyo. Dr. Delgado, your unwavering support and encouragement have been instrumental in preparing me for my academic and professional endeavors. Thank you for consistently challenging me to think critically, facilitating professional networking opportunities, and modeling professionalism. Your mentorship serves as an example that I eagerly anticipate fostering with my future students. Dr. Karen B. Chen, your willingness to invest your time and expertise in my development have truly made a difference in my academic pursuits. I am grateful for the wisdom and dedication you have demonstrated guiding me through the various stages of my doctoral journey.

I am grateful for the invaluable support and mentorship provided by my committee members, including Dr. M. Gail Jones, who provided profound feedback on my work, and Dr. Maria Oliver-Hoyo, whose support and encouragement were indispensable to my growth as a scholar.

I have had the privilege of collaborating with several other faculty, including Dr. Margaret Blanchard, who mentored me during my initial appointment as a graduate teaching assistant, and Dr. Matthew Peterson, who imparted invaluable grant-writing experience. I want to express my appreciation for all the support I have received from my classmates who joined me on this journey and made this process enjoyable, thank you.

A special thanks to my collaborating teacher, for their friendship, support, and allocating instructional time for me to conduct this dissertation research. Finally, thank you to the students who participated in this study. Without your participation, there would be no dissertation to share.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction

Overview

Scientific discoveries at the extremes of scale impact society across multiple levels and within various scientific disciplines (Winkelmann & Bhushan, 2016; National Aeronautics and Space Administration, 2022), establishing the need for a future workforce and citizenry able to navigate across scales. “Scale, proportion, and quantity” (SPQ) reasoning is identified by the Next Generation Science Standards (NGSS) as a crosscutting concept that pervades science, acting as the bridge between scientific disciplines to establish coherent integrated scientific knowledge (NGSS Lead States, 2013). However, research has demonstrated students’ and teachers’ difficulty navigating such extremes of scale (e.g., Tretter et al., 2006b; Jones et al., 2008), which might be attributed to their abstractness (Sabelli et al., 2005) or the inability to experience entities at the extremes of scale (Magaña, 2014). However, instructional interventions to date have only demonstrated a limited ability to support students’ conceptions of size and scale (e.g., Magaña, 2014), have been limited in assessment scope (e.g., Chesnutt et al., 2018) or have required a large amount of resources and time (e.g., Delgado et al., 2015). The theory of embodied cognition posits that there is an inextricable link between the environment, mind, and body, which suggests that experiences that foster such connections will support student learning (Wilson, 2002). However, entities (i.e., atoms, molecules, Earth) referenced throughout the NGSS cannot be experienced directly due to their immense size relative to human scale or too small to see. Therefore, virtual reality (VR) is situated to support students’ size and scale reasoning by providing students with direct, embodied experiences of entities at the extremes of scale. However, to date, research exploring how and when VR supports learning is limited (Luo et al., 2021), especially regarding science instruction at the middle school level (Matovu et al., 2023). None of the identified studies specifically explored SPQ. Therefore, this dissertation explored these gaps in the literature, guided by the theory of embodied cognition and prior research into size and scale reasoning suggesting that VR might be a uniquely powerful tool to support student learning. This dissertation sheds light on unresolved issues in learning about size and scale with VR at the middle school level by studying the use and impact of Scale Worlds (SW), an immersive virtual environment developed to support students' size and scale reasoning. SW was incorporated within a nine-day co-designed instructional unit integrating size and scale reasoning with human impacts on Earth systems in a middle school agricultural science class. Using an experimental explanatory mixed methods design (Creswell & Plano Clark, 2018), I explored the impact of SW on 1) students' conceptions of size and scale, and 2) students' integration across SPQ and multiple DCIs. The focus on integration across a CCC and DCIs is important because the NGSS posit that the integration of CCCs with DCIs and SEPs supports students' understanding of phenomena across disciplines, creating a coherent body of scientific knowledge (Nordine & Lee, 2021). It was hypothesized that the direct, embodied experiences of scale, aided by SW, would facilitate the interaction between the environment, mind, and body, resulting in improved students' size and scale reasoning. Furthermore, by improving students' size and scale reasoning, a crosscutting concept, I hypothesized that I would observe an increased integrated understanding of the DCIs (e.g., human impacts on Earth systems) guiding instruction among the group of students' utilizing SW (NGSS Lead States, 2013). This dissertation produced a VR-based instructional intervention that integrates a CCC and DCI for use within real educational contexts, expands the research regarding VR inclusive science instruction, and contributes to the field's research on the interactions between CCCs and DCIs.

Statement of the Problem

Technological advances have moved our control of matter from the macroscale to the atomic, requiring professionals to navigate scale worlds far beyond everyday experience (Winkelmann & Bhushan, 2016). For example, nanotechnology, applications in the size range from 1 - 100 nm, have led to global advances in medicine, energy, and communication (National Academies of Sciences, Engineering, and Medicine, 2020). Advances in nanotechnology will provide clean water, greener technologies, and novel medical treatments (National Nanotechnology Initiative, 2021). Nanotechnology's applications in agricultural science include the creation of nano-based agrochemicals for increased crop yield and plant growth (Khan et al., 2022) and the development of nanosensors to detect food spoilage to improve food delivery (Ashfaq et al., 2022) or efficiently detect crop needs (Singh et al., 2022). At the other extreme scale, we continue to unlock the secrets of the universe, imaging the black hole at the center of our galaxy, which was precisely the size predicted by Albert Einstein's general theory of relativity (Castelvecchi, 2022).

Understanding the scientific account of Earth's global systems, a necessity for addressing our ever-changing climate and improving quality of life, requires knowledge across scales (National Aeronautics and Space Administration, 2022; Mohan et al., 2009). Therefore, we need a future workforce able to navigate across scales as we continue to establish new cornerstones of science. For multiple professions across domains, scale is seen not just as important but an essential skill to accomplish work-related tasks (Jones & Taylor, 2009). To support and challenge novel science, we must aid our citizens in navigating across scales to make informed decisions regarding the future of science.

Scientific knowledge is not a composition of isolated ideas but often connected in context as students wonder and explore their world. Curiosity motivates students to apply these ideas across contexts, continuing to establish links and modify their ideas (Linn & Eylon, 2011). Despite students' preference to learn science by doing (Corliss & Spitulnik, 2008), a majority of science instruction employs the absorption approach, or the delivery of knowledge (Linn & Eylon, 2011). This disservice results in students' unaware of the complexity of science and unprepared to tackle the scientific problems of the future (National Science Foundation, 2020). Linn & Eylon (2011) suggest the knowledge integration approach, which recognizes the diverse ideas about science that students possess prior to instruction, engages and supports students' ability to integrate ideas to prepare them to become active citizens and scientists (Collins & Halverson, 2009).

Relatedly, the Next Generation Science Standards (NGSS) provides a framework for structuring and implementing science education through a three-dimensional approach: crosscutting concepts (CCC); science and engineering practices (SEP); and disciplinary core ideas (DCI; NGSS Lead States, 2013). The NGSS identifies "scale, proportion, and quantity" as a crosscutting concept that pervades science, a notion supported by Sakhini and Blonder (2018), who recommend the integration of essential nanoscale science technology concepts such as size and scale reasoning within K-12 science curriculum. Scientific ideas do not operate in isolation; crosscutting concepts, such as SPQ, are common across the disciplines of science and engineering (Nordine & Lee, 2021). For example, knowing the strength of hydrogen bonds alone is not useful for guiding an investigation into a basilisk's ability to walk on water; students will also need to consider the scale of the organism and the strength of water's surface tension, demonstrating how CCCs and DCIs interact in supporting a scientific practice (designing an investigation). Exploring phenomena, a central cornerstone of science, thus requires the integration of the three dimensions of science learning (Nordine & Lee, 2021).

A review of the NGSS and College Board Readiness science standards (NGSS Lead States, 2013; College Board, 2009) revealed that 23% of entities referenced were at the atomic scale (Gampp et al., 2022). This presents a challenge given that previous findings have shown students' difficulty navigating this scale with some considering atoms macroscopic (Griffiths & Preston, 1992). In addition, students have been shown to confuse the atomic scale with the microscopic scale, considering atoms and cells similar in size (e.g., Flores et al., 2003). Discerning the size and scale of entities at the extremes of scale presents a challenge for learners of various ages, including middle school students (e.g., Delgado et al., 2015), high school students (e.g., Tretter et al., 2006a), undergraduate students (e.g., Magaña et al., 2012a), and teachers (e.g., Jones et al., 2008). For example, Castellini et al. (2007) demonstrated that only 7% of people surveyed from 7 to 91 years of age were able to correctly rank the relative sizes of atom, water molecule, and cell, which are some of the most frequently referenced entities within the NGSS (Gampp et al., 2022). Tretter et al. (2006b) found a precipitous drop in accuracy of students' conceptions for objects too small to see for various learners (i.e., elementary, middle school, high school, gifted, and expert), which Jones et al. (2008) observed among novice and experienced teachers.

Learners' difficulty might be attributed to the abstractness of entities at the extremes of scale (Sabelli et al., 2005). Paivio's (1990) dual coding theory of cognition proposes that we process non-verbal phenomena and language through separate subsystems. They hypothesized a discrete processing ability that can only encompass sufficiently repeated experiences (Paivo, 1990). Therefore, size and scale concepts may be limited due to students' scarce experiences with objects too small to see or too large relative to human scale (Tretter et al., 2006a). To address this, researchers have been designing new ways to interact with this unseen world since the 1960s, from 2D models to 3D digital manipulatives (Lancaster, 2018). In 1968, Charles Eames developed “Powers of Ten,” a film that illustrates the logarithmic nature of scale zooming in from the universe to a carbon atom. Jones et al. (2007) found that after watching the 21-minute video middle school students significantly improved in their ability to order objects by size, indicating additional research on instructional tools is needed, especially those incorporating increased interactivity. Magaña (2014) compared the exploration of size and scale through three multimedia tools among undergraduate students and found that learning strategies that promote object comparisons might support students’ ability to order and group objects by size. Yet, these multimedia tools have only acted as windows into such worlds. Present instructional tools that support size and scale reasoning have been limited to the desktop with minimal interactivity, not involving full body interactions with entities at the extremes of scale, which Magaña (2014) suggests induces difficulty. However, virtual reality’s ability to facilitate experiences not possible in reality may be powerful for learning about size and scale.

Immersive virtual reality (VR) affords educators the impossible; allowing perplexing and alien concepts to be translated into adaptable, sensory-rich experiences that have been shown to enhance student learning (Luo et al., 2021). Steffen et al. (2019) posit that a primary advantage of VR is its ability to establish a sense of presence or a sense of “realness” (Lee, 2004), within highly adaptable environments (i.e., not limited by physical world constraints). Wilson (2002) posits that human thought is not merely an abstract process contained within the mind but deeply rooted in our environmental interactions, proposing that the environment and our body are part of our cognitive system. Therefore, there is an inextricable relationship between the mind and the body as we interact with the world, which promotes knowledge acquisition. VR provides educators with novel instructional opportunities to elevate abstract instruction through embodied interactions that are within economic reach (Luo et al., 2021). Experiencing scientific entities within VR facilitates embodied cognition by allowing students to interact with 3D scientific entities, which Magaña (2014) suggests might further support their ability to order and group objects by size.

To date, there is limited research on VR in K-12 education (Luo et al., 2021). A review of the papers on VR designed, implemented, and evaluated for science education, showed there is little in middle school, with 58% of papers at the post-secondary level, 9% at the high school level, and 9% at the elementary level with only 10% of papers focused at the middle school level (Matovu et al., 2023). A majority have focused on activities outside the typical classroom or curriculum indicating the need for further investigation in real educational contexts (Matovu et al., 2023). For example, of the studies identified by Matovu et al. (2023), only Liu et al. (2020) investigated the use of VR to support students' conceptual understanding of the characteristics of animals and plants within a physical middle school classroom that significantly improved students affective and cognitive outcomes, highlighting the possibilities of VR instruction. A review of the studies identified by Matovu et al. (2023), searching for references to the NGSS or crosscutting concepts, demonstrated a gap in literature regarding VR and the integration of DCIs and CCCs. However, multiple studies have exploited the affordances of VR to support students' conceptions of quantity (e.g., Hsu et al., 2018), the relationship between scale and function (e.g., Gochman et al., 2019), and entities at the extremes of scale (e.g., Bibic et al., 2019). For example, Ferrell et al. (2019) developed a virtual environment that illustrated the movement of a methane molecule through a carbon nanotube to engage the students with the microscopic world. In addition, present research has primarily assessed learners' recall knowledge, indicating a need for VR studies exploring additional knowledge types (Matovu et al., 2023). To date, there is no consensus on VR's impact on learning outcomes. When comparing VR to alternative learning media (e.g., desktop computer, direct instruction approaches), VR was only more effective in 55% of studies and less effective in 12% of studies (Matovu et al., 2023). However, 82% of interventions evaluated through pre- and post-tests without a control documented increased science learning outcomes, showing the promise of VR in education (Matovu et al., 2023). For example, Maksimenko et al. (2021) found that students' ability to create atoms from subatomic particles in VR during a single undergraduate class period improved students' understanding of atomic structure. However, when comparing the use of slides against the use of VR for teaching scientific knowledge, Parong and Mayer (2018) found that reviewing slides led to significantly better performance compared to the use of VR.

This is not to suggest that VR negatively affects learning outcomes but to instantiate the need for further research exploring how and when VR is more effective than traditional media.

Purpose

Given the promise of and unsettled questions concerning VR to learn about scale, this dissertation investigates the implementation of an instructional unit that incorporates Scale Worlds (SW), a virtual environment, where students can shrink to the size of an atom, grow to the size of the Sun, and explore orders-of-magnitude in between. SW's development was theoretically aligned with contemporary research on size and scale cognition (Delgado, 2009; Tretter et al., 2006a, 2006b; Magaña et al., 2012a) while incorporating educational multimedia design principles (Mayer, 2001) and usability standards (e.g., Nielsen, 1994). For example, SW consists of a series of scale worlds at each order-of-magnitude from 10^{-10} m to 10^{9} m based on experts need to "mentally jump to a new world" to conceptualize the size and scale of various entities (Tretter et al., 2006b, p. 1077). Learners begin the experience in the human world to establish "at least a tenuous link" to everyday scale, which experts described as necessary to process and connect multiple scale worlds (Tretter et al., 2006b, p. 1081). To create substantial connections with science education, entities and boundary limits for the environment were identified through a document analysis of national science education standards (Gampp et al., 2022). Supports and interactions within the environment were developed to elicit the five fundamental ways of thinking about size and scale identified in Magaña et al.'s FS2C (2012a). For example, entities are presented in serial order to foster comparisons across scale worlds, which Magaña (2014) suggests supports students' ability to order and group objects by size. To support students' relative size conceptions and reinforce the logarithmic nature of scale, students can enable a stacking feature where one entity is unitized to measure another entity, a process used by experts for navigating extremes of scale (Tretter et al., 2006a).

SW has been evaluated by usability experts (Wu et al., 2023) and holistically reviewed by middle school students (Harper-Gampp et al., 2023c) prior to implementation. SW was incorporated in a lesson co-designed with middle school science teachers, which produced results descriptively suggesting an improvement of students' size and scale reasoning (Delgado et al., 2023). However, our initial implementation assessed a limited scope of size and scale reasoning, did not include any measures of non-scale science content learning gains, substantiating the need for further research on SW's impact on student cognitive outcomes.

To address these gaps in the literature, this dissertation study implemented an experimental explanatory sequential mixed methods design (Creswell & Plano Clark, 2018) to investigate SW's impact on middle school students' size and scale reasoning, and integration of the crosscutting concept with an understanding of human impacts on Earth systems, a disciplinary core idea. In partnership with an agricultural science teacher from Blue Lake Middle School (pseudonym), a Title I school located in the southeastern United States, I co-developed a unit investigating the socio-scientific issue (SSI) regarding the relationship between cropland and human systems through an iterative process. North Carolina is of particular interest because it is projected to lose 1.2 million acres of agricultural land (11.6%) by 2040 due to urbanization, one of the highest in the country (Xie et al., 2023). Blue Lake Middle School's county alone is projected to lose more than 25% (Xie et al., 2023). Therefore, a nine-day instructional unit exploring the relationship between cropland and human systems within local contexts was co-developed, integrated with SPQ and guided by an emergent framework for SSI education (Sadler, 2011). VR was purposefully integrated within the first three instructional activities and available for use for the last activity with time during the first instructional day devoted to introducing students to the technology to mitigate cognitive demand (Meyer et al., 2019).

Research Questions

This dissertation utilized an experimental explanatory mixed-method design (Creswell and Plano Clark, 2018) with two primary focuses: size and scale reasoning; and knowledge integration. As recommended by Creswell and Plano Clark (2018), I identify a quantitative, qualitative, and mixed methods research question to guide each study.

Study 1: Investigating Middle School Students' Scale Reasoning Through Immersive Virtual Reality-Supported Science Instruction

- How does the size and scale reasoning of middle school students' change after VR-supported science instruction? (QUAN)

- What role does Scale Worlds VR play in students' development of size and scale reasoning? (qual)
- How do middle school students' interviews help to explain quantitatively observed changes in scale reasoning after VR-supported science instruction? (QUAN→qual)

Study 2: Exploring Human Impacts on Earth Systems Across Scale Worlds in Immersive Virtual Reality: An Investigation of Students' Integrated Understanding

- How do middle school students integrate a crosscutting concept and a disciplinary core idea?
- How does immersive virtual reality-supported instruction influence middle school students' integrated understanding of a crosscutting concept and a disciplinary core idea?

Significance

Previous interventions to support students' conceptions of size and scale have had a limited ability to support students' conceptions (e.g., Magaña, 2014), been limited in assessment scope (e.g., Chesnutt et al., 2018) or required a large amount of resources and time (e.g., Delgado et al., 2015). Therefore, this dissertation sought to develop and evaluate a VR-based instructional intervention to explore the use of VR as an instructional tool to support size and scale reasoning based on its ability to simulate embodied experiences at the extremes of scale. This dissertation contributes to the lack of research on when and how to use VR to support middle school science instruction (Matovu et al., 2023) with the goal of determining the efficacy of VR to support students' conceptions of size and scale. Therefore, this dissertation addresses a lack in the types of knowledge explored, with a majority focused on knowledge recall (Matovu et al., 2023), regarding VR inclusive science instruction by investigating students' integration of SPQ, and the DCIs described above.

Framework

Cognition was originally isolated within mind, depicted as abstract interactions guided by mere inputs from experience with our environment (Wilson, 2002). Embodied cognition recognizes the body as inextricably linked to the mind, where cognition resides at the nexus of the environment, body and mind (Clark, 1999). Therefore, learning is more than an isolated individual activity but an environmental interaction. Physical interactions with objects beyond the everyday scale are impossible, but virtual, embodied interactions can be experienced by learners through VR.

Summary of Methodology

Using an experimental explanatory sequential mixed methods design (Creswell & Plano Clark, 2018), I investigated students' scale reasoning using the Assessment of Size and Scale Cognition (ASSC), a previously validated instrument (Harper-Gampp et al., 2023a, 2023b, 2023d), at baseline and conclusion. I investigated students' DCI and SPQ integration through a knowledge integration assessment at baseline and conclusion to address the lack of studies investigating types of knowledge beyond recall knowledge (Matovu et al., 2023). The assessment consisted of ten free response items aligned to instruction and piloted among a small group of the sample population. Survey responses were first coded for the presence of a DCI and SPQ. Following a constant comparative method (Corbin & Strauss, 2008), connections across integrated responses, those coded for both a DCI and SPQ, were inductively analyzed to extract patterns, categories, and themes. Integrated responses were then compared across groups. A subset of participants was invited to participate in a follow up semi-structured interview to elicit individual perspectives and further explore quantitative findings. It was hypothesized that the direct, embodied experiences of scale, facilitated by Scale Worlds VR, would facilitate the interaction between the environment, mind, and body, resulting in improved students' size and scale reasoning. Furthermore, by improving students' size and scale reasoning, a crosscutting concept identified by the NGSS, I hypothesized that I would observe an increased integrated understanding of disciplinary core ideas among the group of students' utilizing Scale Worlds (NGSS Lead States, 2013).

Limitations

This dissertation was isolated to eighth grade students' conceptions of size and scale and might not be generalizable to other crosscutting concepts, science topics, or other grade levels. The novelty of implementing VR within a traditional classroom setting, specifically regarding using it as an instructional tool, introduced challenges that might have made it less effective. Similarly, while time was devoted to supporting students' understanding of how to use VR, some difficulty with the equipment persisted as observed in our previous implementation (Harper-Gampp et al., 2023c). Finally, VR has been shown to induce motion sickness (Caserman et al., 2021), which might have limited some students experience in the VR environment.

CHAPTER 2: Background

Next Generation Science Standards

The National Academy of Sciences developed the Framework for K-12 Science Education, based on the current research, that identified a set of broad expectations for science and engineering K-12 students in the United States conveyed through three dimensions: disciplinary core ideas, crosscutting concepts, and science and engineering practices (NRC, 2012). Students continue to build on disciplinary core ideas, or essential disciplinary concepts, throughout their educational career. The science and engineering practices are the ways in which students engage with the content, seeking to emulate the way scientific knowledge develops. Crosscutting concepts create a coherent conception of science by acting as the bridge between disciplines. Building on the initial framework, the Next Generation Science Standards (NGSS), were developed collaboratively among states and various stakeholders to establish internationally-benchmarked science standards that prepare students for success (NGSS Lead States, 2013). The NGSS proposes a set of student performance expectations, which integrates the three dimensions identified in the initial framework, that are organized within a domain: physical science; life science; earth and space science; and engineering, technology, & applications science. It is through this vertically aligned three-dimensional learning that the NGSS propose students will gain the necessary skills for success in college and 21st century careers (NGSS Lead States, 2013).

Knowledge Integration

Science is about exploring the unknown, expanding the collective body of knowledge. However, a majority of science instruction employs the absorption approach in which students are expected to absorb knowledge through direct instruction, textbooks, and cookbook-like activities (Linn & Eylon, 2011). While this type of instruction supports educators' requests for elegant, well-organized accounts of science, it leads to middle school students believing that memorization is better than understanding science (Madhok, 2006). Knowledge integration (KI) prioritizes the integration of ideas to help learners become productive citizens and scientists (Collins & Halverson, 2009). The KI approach consists of four processes: elicit ideas, add ideas, distinguish ideas, and reflect on ideas (Linn & Eylon, 2011). Initially, KI instruction encourages students to develop ideas about a particular scientific concept that are contextualized within their own personal experiences. They then expand their ideas to incorporate normative ideas regarding the scientific topic of interest. With guidance, the student establishes criteria for distinguishing between their original ideas and the novel ones. Finally, students reflect on their evidence to connect and synthesize a more integrated understanding of the scientific concept (Linn & Eylon, 2011).

Well-known taxonomies of knowledge distinguish among factual, conceptual, procedural, and metacognitive. Factual knowledge is knowing basic elements, bits of information: knowing that a hectare is 10,000 m^2 or knowing that farms grow crops and raise animals. Conceptual is the "interrelationships among the basic elements within a larger structure that enable them to function together (Anderson & Krathwohl, 2001, p. 29)." Therefore, knowing more organized and complex bodies of knowledge (i.e., concepts) is referred to as conceptual knowledge (Anderson & Krathwohl, 2001, p. 29). For example, knowing that hydrogen bonds are strong, a component of a disciplinary core idea within the NGSS, is factual knowledge (NGSS Lead States, 2013). Conceptual knowledge consists of integrated elements organized to explain phenomena. For example, a mental model explaining why a basilisk lizard can walk on water may include ideas about hydrogen bonds, molecular structure, characteristics of water such as surface tension, and animal adaptations. This mental model consists of the integration of disciplinary core ideas (e.g., properties of water) and a crosscutting concept (i.e., scale). However, VR based science instruction research has primarily investigated recall knowledge gains with few studies assessing procedural knowledge (Matovu et al., 2023). Therefore, this dissertation will seek to address this gap by assessing students' integration of disciplinary core ideas with crosscutting concepts to expand present research on integrated knowledge and its development within VR based instruction.

Crosscutting Concepts

Crosscutting concepts that unify scientific disciplines have appeared in standard documents and reports, dating back to Science for All Americans (AAAS, 1989). The project identified four "common themes" that pervade science, mathematics, and engineering: systems; models; constancy and change; and scale. In 1993, the Benchmarks for Science Literacy referenced the same common themes, elaborating on their pervasiveness beyond science, mathematics, and engineering (AAAS). The Framework for K-12 Science Education and the NGSS established the three-dimensional approach, purposefully integrating crosscutting concepts within each performance expectation, the combination of a disciplinary core idea, a science and engineering practice, and a crosscutting concept (Nordine & Lee, 2021). The NGSS identified seven crosscutting concepts: patterns; cause and effect; scale, proportion, and quantity; systems and system models; energy and matter; structure and function; and stability and change (NGSS Lead States, 2013). However, the implicit inclusion of crosscutting concepts in the past has limited students' ability to engage in meaningful science learning (NRC, 1996). They have received relatively little explicit attention within the classroom with their precise role in students' science learning still actively being explored (Nordine & Lee, 2021). Nonetheless, a majority of the community agrees that explicit instruction of crosscutting concepts supports students' learning by increasing connections across disciplines and providing a lens to explore scientific phenomena (Nordine & Lee, 2021).

Scale, Proportion, and Quantity

Initially, it is natural to think of atoms as tiny solid, spherical, and stationary things that make up everything. In fact, many students have difficulty appreciating the constant random motion and emptiness, relying on their everyday experience when developing ideas of atoms (Kind, 2004). Therefore, we must recognize how phenomena change relevant to scale, but also proportion, and quantity. For example, appreciating speed as the ratio of distance traveled to time taken, or grasping the difference between a few parts per million versus a few parts per billion water contaminants. Therefore, the Framework for K-12 Science Education (NRC, 2012), identifies scale, proportion, and quantity (SPQ) as a valuable concept for establishing a cohesive understanding of science across disciplines. For example, chemists navigate the atomic world in search of key characteristics that elevate macroscopic properties. Biologists model cellular activity in search of mechanisms to alter or understand to improve and extend human life. Environmentalists analyze global cycles of precious resources to maintain and support a healthy global ecosystem. Astronomers look beyond our planet to the depths of space to unravel the mysteries of our universe. SPQ acts as the tool that each of these scientists employ to understand such multiscale phenomena. For example, students often confuse cells and atoms (Flores et al., 2003), which might be a result of both objects being described as “building blocks” within their respective disciplines (Delgado et al., 2021). However, if students had a sense of their size then they might more easily differentiate the two, developing a more cohesive conception of science (Delgado et al., 2021). While the NGSS considers SPQ as a single crosscutting concept, a prior operationalization of SPQ to classify assessment items (Laverty et al., 2016) identified scale as sufficiently different from proportion and quantity. Scale was considered as comparisons between objects, processes, or properties across size, time, or energy scales while proportion and quantity considered the response of one variable to changes in another. Therefore, modeling this classification, for this dissertation, SPQ was split into two independent crosscutting concepts: scale (S) and proportion and quantity (PQ).

Size and Scale Reasoning

To obtain a comprehensive understanding of recent research on size and scale reasoning, I systematically surveyed peer-reviewed academic studies published in the period 2013-2023 following the PRISMA methodology (Figure 2.1; Page et al., 2021). An electronic literature search was conducted on two scholarly databases (Web of Science and ERIC). Web of Science was chosen because it consists of a large repository of journal articles from a variety of disciplines. Similarly, ERIC was chosen for its expansive coverage of educational research. Initially, the general search string ‘size’ OR ‘scale’ was employed but returned a large number of studies (>3,000,000). To limit the search to relevant literature regarding size and scale reasoning, foundational papers were mined to identify prominent keywords (e.g., Tretter et al., 2006a, Jones et al., 2007; Delgado, 2009; Magaña et al., 2012a). These papers were identified from the proposal for the National Science Foundation-funded research project that developed Scale Worlds. The final search string employed was ‘size and scale’ OR ‘scale and size’ OR ‘scale cognition’ OR ‘size cognition’ OR ‘scale reasoning’ OR ‘size reasoning’. The initial search yielded a total of 461 documents (Web of Science 435 and ERIC 26). Study titles were then screened for type and field of study. Only peer-reviewed academic journal articles within science education were retained. Studies in fields such as business, natural sciences, and medicine were excluded (e.g., Altin et al., 2017; Aldas & Yapici, 2014; Attoh-Mensah et al., 2020). Based on these criteria, 439 studies were excluded. A review of the remaining studies' abstracts led to the exclusion of 10 additional studies given their lack of focus on size and scale reasoning. For example, Lin et al. (2013) was excluded because the primary focus was on the public's attitudes towards nanoscience not size and scale. A total of 12 studies (Table 2.1) were retained for full review. I discuss these within the context of previous work on size and scale reasoning to establish a clear and cohesive narrative of extant literature.

Figure 2.1 Systematic review flow diagram (adapted from Page et al., 2021).

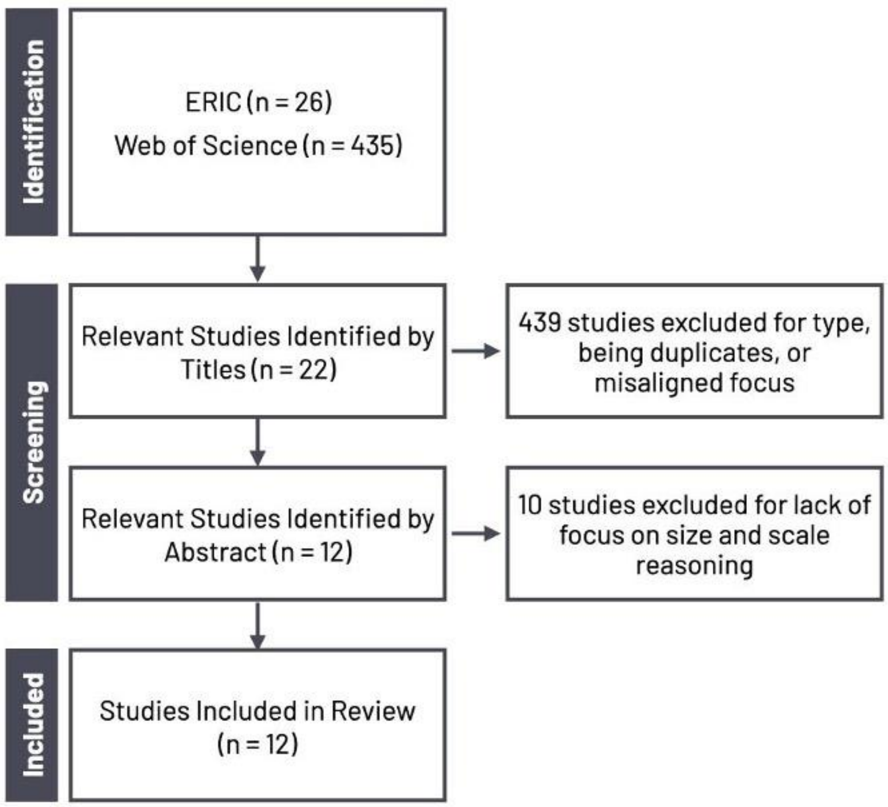


Table 2.1 Articles retained for full review.

Article	Short Description
Chesnutt et al., 2019	Longitudinal analysis exploring the relationship between size and scale reasoning, and students' achievement in science and math. A positive relationship was observed for both.
Chesnutt et al., 2018	Conceptualized science scale capital based on Bourdieu's social capital (e.g., 1977). Race, out-of-school experiences, and exposure were correlated with absolute scale performance.
Delgado et al., 2015	A 12 hr summer camp to support conceptions of size and scale including multiple activities. Public middle school students' performance elevated to private middle school and public high school students' performance.
Gerlach et al., 2014a	Investigating undergraduate students' conceptions of size and scale through activities modeled after previous assessments. Demonstrated that experience did not necessarily indicate improved scale literacy.
Gerlach et al., 2014b	Developed Scale Literacy Skills Test and Scale Concept. Administered to chemistry students of varying experience.
Jones et al., 2013	Compared the size and scale reasoning of teachers from the United States, Taiwan, and Austria. Showed that Austrian and Taiwanese held more accurate conceptions of size and scale.
Kong et al., 2017	Expanded Magaña framework to encompass definition, grouping by scale, and comparing numerical scales.
Magaña, 2014	Compared the use of various multimedia scale tools' ability to support conceptions of size and scale. Suggested tools support students' ability to order and group by size.
Mandrikas et al., 2020	Developed teaching learning sequences on size and scale for 6th grade students focused on nanoscience and nanotechnology concepts. Results demonstrated some students' decompression of scale.
Resnick et al., 2017	Identified a framework for how relational reasoning bridges everyday experiences to the extremes of scale, identified barriers, and outlined instructional methods to teach scale.
Trate et al., 2019	Replicated investigation of size and scale among undergraduate chemistry students. Supported previous findings that students demonstrate increased accuracy at scales close to human scale.
Vlaardingerbroek et al., 2014	Describes misconceptions induced by illustrations not depicting the true relative size of cellular components.

Frameworks for Size and Scale

While scale has been identified as a common theme since the release of Science for All Americans (AAAS, 1989), research on scale did not emerge until 2006a with Tretter et al's analysis of conceptual boundaries being one of earliest studies on learners' conceptions of size and scale. In 2007, Light et al. captured 12 undergraduate students' various ideas of size and scale through qualitative interviews, demonstrating variability in students' conceptions. Based on the students' conceptions, Light et al. (2007) developed an initial typology to capture the increasing complexity of students' responses. They proposed that size refers to the actual amount of something, while scale links the phenomenon to a numerical representation of size. Following the work of Light et al. (2007), Swarat et al. (2011) extended this exploration while considering size and scale as a unifying idea among undergraduate students. The authors developed a typology consisting of four conception categories (fragmented, linear, proportional, and logarithmic) with seven subcategories to account for subtle differences in examples. To establish a framework for learning, Delgado (2009) utilized student interviews to develop a learning progression on how middle school, high school, and college students make connections across four aspects of size and scale identified from Piaget's work: ordering (seriation), grouping (classification), relative size, and absolute size, where the first two are qualitative and the last two quantitative. Building on much of this earlier work, Jones and Taylor (2009), through interviews with professionals, developed a trajectory of scale development from more abstract to concrete conceptions.

Magaña et al. (2012a), also informed by Inhelder and Piaget (1958), developed a knowledge framework to categorize and scaffold size and scale cognition (FS2C; Table 2.2) that guides this dissertation. The FS2C consists of five fundamental ways of thinking about size and scale: qualitative relational conception (ordering); qualitative categorical conception (grouping); qualitative proportional conception (logical proportional reasoning); quantitative proportional conception (relative size); and quantitative absolute conception (absolute size). Qualitative relational conception and qualitative categorical conception are modeled on the processes of discrimination and generalization. Discrimination consists of the seriation (serial ordering) of objects by size (atom < virus < white blood cell < human) while generalization is the classification of objects relative to categories of scale (e.g., atomic, nano, micro, macro). These qualitative conceptions of size are linked to quantitative conceptions of scale through proportional reasoning. Inhelder and Piaget (1958) posit that proportional reasoning is composed of a logical and a mathematical component. The logical form establishes a relationship between a set of objects (A & B) and another set of objects (C & D), suggesting that the logical component preludes the mathematical component. The mathematical form of proportional reasoning consists of assigning a numerical value to represent quantity: object B is 100 times smaller than object A. The final way of thinking of size and scale is the quantitative absolute conception, or assigning a numerical measurement to an object. Delgado and Peterson (2018) identified a specific instance of logical proportional reasoning, nested logical proportional reasoning, where within the equivalence ratio each set shares a common object (i.e., A is to B as B is to C; where B is the common object among both sets), and posited that this instance of logical proportional reasoning might be a natural and intuitive way to learn based on humans' use of conceptual metaphor to create meaning by drawing connection between a known source concept and a foreign target concept (Lakoff & Johnson, 1980). Based on this previous work, I adopted the Magaña FS2C framework with the addition of NLPR developed by Delgado and Peterson - see Table 2.2 below.

Kong et al. (2017) sought to extend the work of Magaña et al. (2012a) by analyzing undergraduate engineering students' conceptions of size and scale through a qualitative analysis of their nanotechnology-based projects. They recommend expanding Magaña et al.'s (2012a) FS2C to include definition, quantitative categorical, and quantitative relational, with the latter two referring to explicitly quantitative strategies for ordering and grouping. Following Delgado (2009), we consider the approaches strategies rather than fundamental conceptions. Using their expanded framework, they developed and validated the Size and Scale Framework-based quiz to assess size and scale reasoning within the context of nanoscience. For the scope of this dissertation, I retained the original framework proposed by Magaña et al. (2012a), including NLPR as developed by Delgado and Peterson (2018), given their alignment with Scale Worlds and the assessment instrument.

Table 2.2 Adopted size and scale reasoning framework (Magaña et al., 2012a; Delgado & Peterson, 2018)

Cognitive Process	FS2C Framework	Description
Discrimination	Qualitative relation conception	Ordering a set of entities by size
Generalization	Qualitative categorical conception	Grouping or categorizing entities by size
Mathematical reasoning	Quantitative absolute conception	Identifying the absolute size of the entity
Numerical proportional reasoning	Quantitative proportional conception	Comparing the relative size of entities
Logical proportional reasoning	Qualitative proportional conception	Comparison of a pair of entities with another pair of entities of an equivalent size ratio (e.g., A is to B as C is to D)
Nested logical proportional reasoning		Comparison in which within the equivalence ratio each set shared a common entity (e.g., A is to B as B is to D)

Prior Research on Student Knowledge and Learning of Size and Scale

Extant literature (e.g., Tretter et al., 2006a, 2006b; Jones et al., 2008) has demonstrated students' difficulty discerning the size and scale of multiple objects. Tretter et al. (2006a) administered an assessment, a card sorting activity and conducted individual interviews among 215 participants (grades 5, 7, 9, and 12, and doctoral students) to identify students' cognitive frameworks and their development over time. They found that experts held more distinct categories at the small scale, with experts integrating more specific experiences compared to novices. Among all participants, relative scale reports were more accurate than absolute scale. Gerlach et al. (2014a) demonstrated that, while graduate level chemists have an increased awareness of sizes below what the eye can see, the increase in experience did not determine high scale literacy (Gerlach et al., 2014a). Similarly, Trate et al. (2019) showed that undergraduate students in chemistry had difficulty classifying objects as either atomic or macroscopic, not demonstrating developed size and scale reasoning upon entering undergraduate chemistry. Tretter et al. (2006b) found that these absolute scale inaccuracies persist among novices and experts, demonstrating that students' accuracy of absolute scale is not symmetrical. Large scale accuracy declines gradually as scale increases while small scale accuracy drops sharply at scales too small to see. Similarly, Delgado (2009) demonstrated that students' estimates of absolute scale tended to be vast but were narrower for objects close to human scale. Resnick et al. (2016) suggests that the incorporation of relational reasoning, the use of comparisons (both similarities and differences), in instructional techniques might support this sense of discontinuity between human scale and the extremes of scale.

Jones et al. (2007) found that both experienced and novice teachers have more accurate conceptions of large entities compared to small entities, which the author suggests might be due to an individual's ability to see very large entities but not very small entities. To obtain an international perspective, Jones et al. (2013) compared the size and scale reasoning of secondary science teachers in the United States, Taiwan, and Austria. They demonstrated that teachers in the United States possessed less accurate conceptions of size and scale, which might be attributed to the limited use of metric units within American society, suggesting the presence of a sociocultural component (Jones et al., 2013). Similarly, Delgado (2013b) found that students who grew up in a country using the SI ("metric") system performed at a significantly higher level estimating the length of a meter and a conceptual task (coordinating relative size and absolute size) compared to their peers who grew up using US customary ("imperial") units. Building on Bourdieu's idea of social capital (e.g., 1977, 1984), Chesnutt et al. (2018) describes this sociocultural component as science scale capital, which encompasses the experiences, resources, and tools that supports students' conceptions of size and scale. They demonstrated that a students' racial/ethnic identities, experiences with scale outside of school, and exposure to size and scale instruction play a statistically significant role in predicting students' size and scale reasoning. Therefore, we require additional research regarding the cultural component influencing students' size and scale reasoning.

Inaccuracies of scientific entities' size and scale persist among learners of all ages. Castellini et al. (2007) demonstrated that only 7% of people surveyed from 7 to 91 years of age were able to correctly rank the relative sizes of an atom, water molecule, and cell, which are some of the most frequently referenced entities within the NGSS (Gampp et al., 2022). When asked to identify the smallest thing they can think of, middle school students often identify small macroscopic objects (e.g., grain of salt; Waldron, Spencer, and Batt, 2006; Delgado, 2009). Magaña et al. (2012a) corroborated this by finding that students have difficulty discerning the relative sizes of microscale and nanoscale referenced objects. In addition, research has demonstrated that students frequently believe atoms and cells are similar in size (e.g., Flores et al., 2003), or macroscopic (e.g., Cokelez, 2012; Griffiths & Preston, 1992). Brook et al. (1984) observed students ascribing macroscopic properties to particles suggesting that they melted or died while describing the melting of ice. Students' misconceptions regarding basic chemical ideas have been documented across the literature and compiled by Kind, 2004. While students might have naive conceptions, Delgado (2009) showed they tend to recall atoms and cells, suggesting they are potential landmarks for students. This might be attributed to the way in which scientific entities are portrayed to students. For example, Vlaardingerbroek et al. (2014) found that many students thought the components of a cell were depicted in instructional material at an accurate scale relative to the cell in which they were contained. This suggests the importance of providing students with experiences in which true scale is maintained to build upon students' naive conceptions, the goal of Scale Worlds.

Students' conception of size and scale are not only considered a crosscutting concept but a predictor of academic performance. It has been demonstrated that students' conceptions of size and scale are positively correlated to academic achievement in science and mathematics at the middle school level (Chesnutt et al., 2019). This evidence supports the notion of three-dimensional pedagogy and crosscutting concepts' ability to pervade disciplines as described in the NGSS (Chesnutt et al., 2019).

Landmarks

Landmarks (also called benchmarks, or anchors) are reference points we use to make sense of our world, taking a known value and adjusting accordingly (Tversky and Kahneman, 1974). For example, students might use their size, distances on athletic fields, or driving distances as meaningful scale landmarks (Tretter et al., 2006a). When approaching inconvenient units or extremes of scale, many experts utilized a unitizing strategy, creating a new unit for use (Tretter et al., 2006a). As one expert explained, "we could wrap our brain around those numbers [used to measure atoms] in Angstroms as opposed to other units (Tretter et al., 2006a, p. 302)." Similarly, Jones and Taylor (2009) identified various landmarks professionals utilize in navigating scale within their profession. Tretter et al. (2006b) depicted this as experts navigating "worlds" based on their descriptions of the scale of entities at the extremes. We incorporate landmarks across disciplines for sense making. Landmarks improve our estimation of the dates of personal events (Loftus and Marburger, 1983) and improve our ability to estimate data such as population sizes (Brown and Siegler, 2001). For temporal scale, Delgado (2013a) defines *individual landmarks* as those developed through personal experience, and *collective landmarks* as those that a group of students have established, and that the instructor can then rely on for instruction, in the context of deep time. For an event to be considered as a collective landmark, the 25^{th} and 75^{th} percentiles of the estimates of a group of learners must be within an order-of-magnitude of each other and the accepted value must fall within the 25^{th} and 75^{th} percentile estimates. The development of collective landmarks meaningfully scaffolds students' learning of scientific phenomena (Delgado, 2013a). The analysis of the 25^{th} and 75^{th} percentiles estimates can also assist instructors in identifying alternative ideas or events in which student estimates lack clarity. Events in which the accepted value fell within the 25^{th} and 75^{th} percentile but the 25^{th} and 75^{th} percentile estimates were not within an order-of-magnitude are considered accurate but not precise, indicating a lack of knowledge about the age of the event. Events in which the accepted value did not fall within the 25^{th} and 75^{th} percentile but the 25^{th} and 75^{th} percentile estimates were within an order-of-magnitude indicate a widely held alternative idea that instruction can explicitly address. However, there is no prior research on individual or collective landmarks for size (Delgado, 2013a). As a part of the unit, I supported students in the development of accurate landmarks for size that I hypothesized would support their integrated knowledge of human impacts on Earth systems.

Assessment Techniques

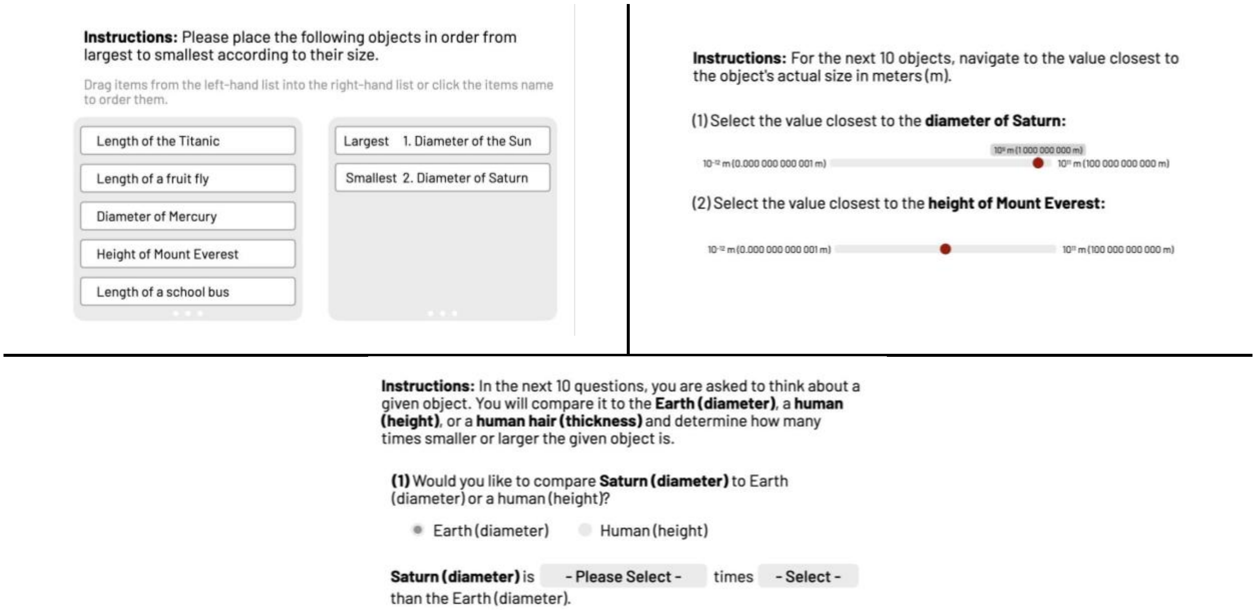
To assess students' size and scale reasoning, several investigations have implemented a physical card sorting activity as a way to gauge students' conceptions of size and scale (e.g., Delgado et al., 2015; Jones et al., 2008; Delgado, 2009; Jones et al., 2007; Tretter et al., 2006a). However, these assessments have consisted of more qualitative analyses (e.g., Delgado et al.,

2015) or the creation of distance matrices for an entire group of students at a given grade level, not individuals, (e.g., Tretter et al., 2006a) with limited discussions of validity and reliability. To assess the generalizability of previous findings on undergraduate chemistry students' conceptions of size and scale, Trate et al. (2019) incorporated a card storing and grouping task as a laboratory activity. They demonstrated that with proper training of the teaching assistants, they were able to acquire results similar to those obtained through individual interviews. However, card sorting assessments are time-intensive and not easily implemented among large groups of students such as a class of students, as demonstrated by Tate et al. (2019).

To assess students' ability to identify the absolute size of multiple entities, Tretter et al. (2006a) developed the Scale of Object Questionnaire (SOQ) based on Trend's (2001) instrument assessing geologic time. The SOQ was reviewed by a panel of science educators and scientists to establish content validity as described by Franklin et al. (2001) and pilot-tested among high school and fourth-grade students to improve clarity. In addition, the Scale Anchoring Objects assessment (SAO), developed by Tretter et al. (2006b), required students to identify entities at discrete sizes smaller and larger than a human. Both the SOQ and SAO have been deployed in multiple empirical studies regarding students' conceptions of size and scale (e.g., Jones et al., 2013; Jones et al., 2008). However, each assessment instrument centered on a single concept of size and scale with no measure of consistency across all five fundamental ways of thinking about size and scale as defined by FS2C (Magaña et al., 2012a).

Magaña et al. (2012a) developed an assessment instrument aligned to each concept identified in the FS2C, limited to a single task item per dimension. Steps to achieve evidence of validity consisted of a comparison to previously established instruments (Delgado et al., 2007; Tretter et al., 2006a), expert review, and independently categorizing each task according to the FS2C. Evidence of reliability was established through a pilot test among middle-school students to assess readability and applicability. Gerlach et al. (2014b) developed and validated the Scale Literacy Test (SLT) and Scale Concept Inventory (SCI) for class wide implementation to explore the scale literacy of college chemistry students. They demonstrated that the combined score of the SLT and SCI, which they refer to as the scale literacy score, was the best predictor of students' success on conceptual test items (Gerlach et al., 2014b).

Figure 2.2 Item examples from each section of the ASSC.



Based on the previously identified limitations, the *Scale Worlds* research project (on which I was a research assistant) developed the Assessment of Size and Scale Cognition (ASSC), a computer-based assessment aligned with the framework to characterize and scaffold size and scale cognition (FS2C; Magaña et al., 2012a). The ASSC includes a task for three of the five fundamental ways of thinking of size and scale identified by Magaña et al. (2012a; see Figure 2.1), assessing student conceptions of ordering, relative size, and absolute size using 20 entities referenced in the NGSS: one for each power of ten from atom to Sun. Face and content validity were assessed through a panel of experts following the methodology of Elangovan and Sundaravel (Figure 2.2; 2021) with some additional processes described in Harper-Gampp et al. (2023a, 2023b, 2023d). After iterative revisions, cognitive interviews were conducted with members of the target population (Harper-Gampp et al., 2023d), leading to a few further modifications. The ASSC was pilot tested among first-year undergraduate engineering students (n = 267) and found to have adequate reliability (Cronbach's alpha = 0.83). The development process ensured strong construct validity.

Table 2.3 Representative entities included in the ASSC.

Size	Entity	Size	Entity
10^{-10} m	Length of a carbon dioxide molecule	10^{0} m	Height of an average adult human
10^{-9} m	Length of a glucose molecule	10^{1} m	Length of a school bus
10^{-8} m	Length of a ribosome	10^{2} m	Length of the Titanic
10^{-7} m	Diameter of the influenza virus	10^{3} m	Length of the Golden Gate Bridge
10^{-6} m	Diameter of the nucleus of a cell	10^{4} m	Height of Mount Everest
10^{-5} m	Diameter of a skin cell	10^{5} m	Length of the Panama Canal
10^{-4} m	Diameter of a human egg cell	10^{6} m	Diameter of Mercury
10^{-3} m	Length of a fruit fly	10^{7} m	Diameter of Venus
10^{-2} m	Length of a bumblebee	10^{8} m	Diameter of Saturn
10^{-1} m	Height of a squirrel	10^{9} m	Diameter of the Sun

Interventions

While size and scale have been identified as a crosscutting concept, there is little evidence that it is addressed during classroom instruction, substantiating the need for instructional tools, resources, and support (Gerlach et al., 2014a; Jones & Taylor, 2009). To support students' size and scale reasoning, Charles Eames developed "Powers of Ten," a film that illustrates the logarithmic nature of scale by moving from a man enjoying a picnic in Chicago to the edge of the universe and to the carbon atoms that makeup his hand (1968). The film was later reimaged for IMAX, a high-resolution movie theater that incorporates large screens, as the "Cosmic Voyage" by the Smithsonian Institution as an exhibit at the National Air and Space Museum (National Air and Space Museum, 1986). While Eames's film was utilized by science educators as a way to promote size and scale reasoning, no research had determined the film's instructional efficacy (Jones et al., 2007). Therefore, using a pre- and post-interview design, Jones et al. (2007) researched the effect that the film had, finding a positive influence on students' conceptions of size and scale, improving their ability to order entities by size. However, students' reported difficulty conceptualizing entities too small to see, suggesting that further research investigate more interactive instructional support (Jones et al., 2007). Jones et al. (2007) indicates that it is imperative to explore the efficacy of various instructional techniques, substantiating the need for exploring how VR might support conceptions of size and scale.

Another approach to support students' size and scale reasoning, was designed, implemented, and evaluated by Delgado et al. (2015). The 12-hour project-based instructional unit was contextualized around the driving question "how can nanotechnology keep me from getting sick?" Students participated in multiple instructional activities of varying modalities (e.g, using a microscope, exploring custom-made computer simulations). Following the intervention, students' overall performance improved to a statistically significant degree, with a large effect size of 0.8. Results revealed the intervention improved students' qualitative and quantitative knowledge of size and scale, closing the gap between the public and private middle school students' performance. In addition, the middle school students overall scored similarly, and descriptively surpassed, public high schoolers, demonstrating several years' worth of learning gains. However, the instructional unit requires a large amount of time and resources, making it difficult to incorporate into an already overwhelmed science curriculum. Even with the inclusion of unitization activities (i.e., measuring objects with invented units) all campers demonstrated difficulty estimating relative size, on average, able to estimate within one order-of-magnitude just 1.5 out of four objects (Delgado et al., 2015).

Building on prior work, Chesnutt et al. (2018) developed a series of activities that students completed during 6th, 7th, and 8th grade to improve students' conceptions of size and scale. Activities were based on previous identified interventions that found significant learning gains. For example, students watched Powers of Ten (Jones et al., 2007) and practiced unitizing by using their leg span to measure distances (Tretter et al., 2006b). The authors found that participation in the intervention significantly improved students' ability to assess absolute scale, illustrating the promise of instructional support. However, the instructional activities were only assessed utilizing the Scale of Objects Questionnaire, which only assesses a single fundamental conception of size and scale, quantitative absolute conception.

Given the salience of size and scale reasoning within nanotechnology, Madrikas et al. (2020) developed a 6th grade teaching learning sequence aimed at improving students' basic knowledge of nanoscience and nanotechnology. Their teaching learning sequence was focused on size and scale, and co-designed with teachers, scientists, science educators, and science center experts. The lesson consisted of discussions of nanoscale topics, video materials, manipulatives, and teleconferences with scientists. Results indicated that some students ' conceptions of size and scale were expanded, indicating that "I learnt some smaller units than centimeters and millimeters called nanometers (p. 387)." However, students' naive conceptions continued to persist with 17% of students' statements suggesting that all non-visible objects are similar in size (Madrikas et al., 2020).

A few interactive web-based multimedia tools have been developed, including Generation-nano (Network for Computational Nanotechnology, 2007), Universcale (Nikon, 2009), and Nanoscaling (Playgen, 2009). Each tool affords users the ability to navigate through various visual images of entities and text descriptions at various sizes, including a logarithmic scale indicating the size of the entity. An evaluation of the three multimedia tools by teachers and instructional designers suggests that a user's ability to compare entities of different sizes is advantageous (Magaña et al., 2012b). However, the incorporation of the tool in teaching and learning was not apparent among several preservice teachers, indicating the need for instructional support in incorporating such tools (Magaña et al., 2012b). An analysis of the instructional efficacy of the tools indicated that they might support students' ability to order and group entities with an increase in their numerical proportional and absolute measurement task performance (Magaña, 2014). Magaña (2014) noted that the students' ability to manipulate 3D objects supported object comparison, suggesting that technological affordances of immersive VR might further support students' ability to compare entities of different sizes.

Embodied Cognition

A majority of professionals (76%), who tend to use scale in their everyday workplace (e.g., chemist, chef, forester), attributed physical experiences as contributing to their scale development (Jones & Taylor, 2009). Early views of cognition have depicted the mind as an isolated stage where the body acts only as an understudy (Clark, 1999). However, research has challenged this notion by insinuating that cognition is poised among the environment, body, and mind (Clark, 1999). The body and mind are inextricably linked as they navigate the world (Wilson, 2002). Wilson (2002) claims that meaningful embodied cognition proposes six distinct claims: cognition is situated, cognition is time pressured, we off-load cognitive work onto the environment, the environment is part of the cognitive system, cognition is for action, and off-line cognition is body based (p. 626). The author indicates that each claim is met with varying degrees of controversy but argues that the first three must be partially true. This suggests that learning is not just an isolated incident within the student's mind but can be facilitated through their interaction with their environment. Therefore, experiences at the nexus of the mind, body, and environment will support student learning. Virtual reality has the ability to create sensory-rich experiences in which users can navigate through the use of bodily movements. Unlike previous technology such as computers or tablets, VR simulates a three-dimensional environment in which individuals interact much like everyday experiences without real world limitations. For example, students can hold a virus in the palm of their virtual hand or dive into the cytoplasm of a cell. Therefore, VR is able to simulate interactions at the nexus of the mind, body, and environment beyond the limitations of reality supporting students' conceptions of the abstract. Jones et al. (2008) and Magaña (2014) suggest that technological affordances be exploited to elicit experiences with objects too large to grasp or too small to see, again pointing to VR as a promising tool to support and develop scale reasoning.

Virtual Reality

Virtual reality (VR) is a complex media system that seeks to emulate or extend our physical world through highly adaptable sensory-rich experiences (Makransky & Petersen, 2021). To date, VR can be accessed through various mediums: desktop computers, Cave Automatic Virtual Environments (CAVE), or head-mounted displays (HMDs). The CAVE, introduced in the 1990's, is a system that surrounds a user in physical space (Cruz-Neira et al., 1993). The environment consists of a partially enclosed square theater in which the side walls, floor and sometimes ceiling act as projectable areas, eliciting a more immersive virtual environment that envelops the user (Ermi & Mäyrä, 2005). However, the CAVE has had limited widespread adoption (i.e., limited to museums or university settings) in part due to financial constraints

(Merchant et al., 2014). An HMD, the focus of this work, achieves a higher degree of immersion at considerably lower cost that requires less equipment and provides increased mobility. The setup consists of a goggle-like headset with embedded screens that the user places over their head, covering their eyes. The system utilizes a combination of dynamic tracking coupled with visual and audio cues to create a highly adaptable sensory-rich experience. For the scope of this work, I adopted the language of Makransky & Petersen (2021), referring to immersive virtual reality (VR) as technology that isolates the learner within a virtual environment through the use of HMD technology. Nonetheless, all VR technology seeks to emulate the natural world by establishing a sense of immersion and presence. Immersion is a property of the system (Slater, 2003) while presence is a psychological state experienced by the user (Lee, 2004).

Virtual Reality in Education

The introduction of VR in education began in the early 1990's with projects such as Science Space and Atom World (Youngblut, 1998). However, despite the increased interest in VR experiences, sustained or widespread implementation in K-12 education has yet to be observed. Kavanagh et al. (2017) identified multiple barriers to implementation, including cost, usefulness, and lack of engagement. Therefore, effective and purposeful educational design must be considered when developing and implementing learning experiences with VR. As Mikropoulos and Natsis (2011) states, "the prerequisite for an effective learning environment is its pedagogical approach" (p. 774). Educators must be able to exploit the unique features of VR to develop meaningful learning experiences and facilitate widespread adoption.

While VR continues to gain attention as a promising learning technology (Thilhou, 2020), research on how and in what situations it might improve learning outcomes is still emerging (e.g., Matovu et al., 2023; Luo et al., 2021). VR in education literature has predominantly focused on desktop VR applications (Luo et al., 2021), which is attributed to their economic feasibility in K-12 education (Kavanagh et al., 2017). For example, a meta-analysis by Merchant et al. (2014) found that virtual worlds were effective in improving learning gains. However, the author's definition of virtual worlds consisted of multiple forms (e.g., augmented reality, mixed reality). As described earlier, to date, there is limited research on VR in K-12 education (Luo et al., 2021), with only 11% of studies on science instruction at the middle school level, and a majority (82%) assessing knowledge recall or attitudes and behavioral changes (Matovu et al., 2023). In addition, while 82% of interventions evaluated through pre- and post-tests documented increased science learning outcomes, showing the promise of VR in education, there is still debate within the community on VR's impact on learning (Matovu et al., 2023). Therefore, the applications and impacts of VR within middle school are still emerging. This dissertation aims to contribute to this body of knowledge continuing to explore how and when VR is more effective than traditional mediums.

Integrating a Crosscutting Concept and Disciplinary Core Idea in VR

Forming an integrated understanding of Earth systems is vital to improve life for all people (National Aeronautics and Space Administration, 2022). Navigating globally complex systems requires students to develop a dynamic and cyclic conception of our planet (Kali et al., 2003). Assaraf and Orion (2005) indicated that difficulties were associated with students' ability to extend system borders to reveal hidden dimensions. For example, only 52% of students described human impacts on well water, indicating a fragmented understanding of system interactions (Assaraf and Orion, 2005). Many students tend to focus only on visible aspects of a system, not discussing microscopic or atomic interactions (Hmelo-Silver et al., 2007). On the scale of human systems (e.g., respiratory system) Hmelo et al. (2000), implied that difficulties partly arise because students did not recognize both macroscopic and microscopic operations. Therefore, students require support navigating dynamic systems across multiple scales. This dissertation supports students' understanding of systems (e.g., urbanization, water) across scales as they explore the relationship between these dynamic systems and agricultural cropland. By exploring croplands role in global systems across scales, students will be challenged with investigating a disciplinary core idea (i.e., human impacts on Earth systems, ESS3.C) integrated with the crosscutting concept of scale, which the NGSS posits supports the development of integrated scientific knowledge (NGSS Lead States, 2013), providing students with the necessary tools to challenge and support the future of science as citizens and scientists (Collins & Halverson, 2009).

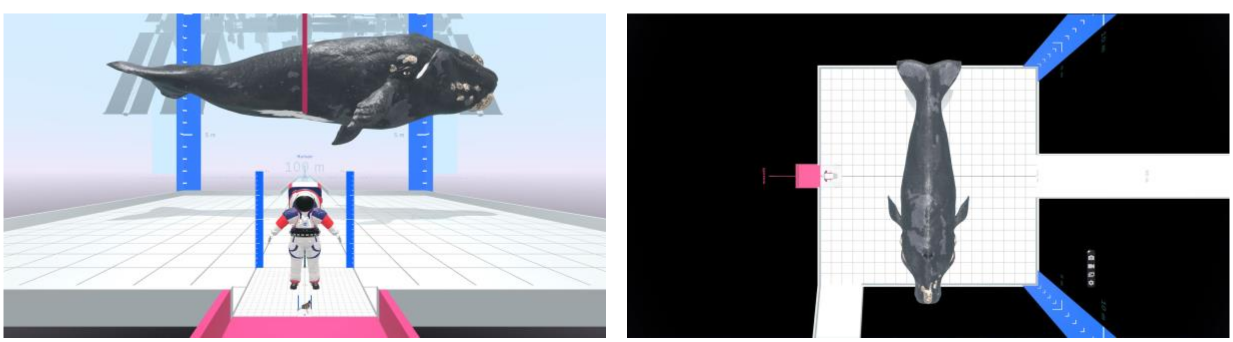
CHAPTER 3: Methods

This dissertation consisted of an experimental explanatory mixed-methods design (Creswell and Plano Clark, 2018) with two primary focuses: size and scale reasoning; and knowledge integration. In this chapter, I outline the methodology, including participant selection and context, description of the virtual environment, data collection techniques, and data analysis procedures regarding each focus.

Context

VR and non-VR participants consisted of eighth grade students' from Blue Lake Middle School (pseudonym), a Title I school located in the southeastern United States. Middle school students were used because of previously established teacher partnerships and the lack of research on VR in middle school (Luo et al., 2021). Eighth grade students are appropriate because Scale Worlds incorporates scientific notation, which is part of eighth grade math curriculum (Common Core State Standards Initiative, 2023), and scientific entities (e.g., atoms, molecules, cells) referenced within middle school NGSS (NGSS Lead States, 2013) and state science standards (North Carolina Department of Public Instruction, 2024). Based on previously established relationships, one teacher at Blue Lake Middle School was invited to participate. This teacher was recruited because they taught agriculture, an elective course, which afforded additional flexibility given the absence of testing pressures. The teacher participated in the iterative co-design and future implementation of a unit on human impacts on Earth systems, guided by the driving question, "how do humans and agriculture interact within our local area?"

Figure 3.1 Scale Worlds-HMD from a user's (A) and a bird's-eye perspective (B).



(A) (B)

Scale Worlds

Scale Worlds (SW) is an immersive virtual environment designed by a National Science Foundation supported project (#2055680). The virtual environment (Figure 3.1) allows students to shrink to the size of a water molecule (10^{-10} m), grow to the size of the Sun (10^{9} m), and explore all ten-fold orders-of-magnitude in between. Versions of SW exist for the Cave Automatic Virtual Environment (CAVE) and head-mounted display (HMD), and a prototype version for web display. For the purpose of this research, I utilized SW-HMD given its developmental state, accessibility, and affordances. The prototype version for web display and Universcale, a previously identified tool that supports size and scale reasoning (Magaña, 2014), were utilized in place of the VR tool for the comparison group.

SW was developed and framed around Magaña et al.'s (2012a) framework to characterize and scaffold size and scale cognition, including the addition of multiple supportive armatures (i.e., vertical measuring devices) and interactions (i.e., entity stacking). We identified entities (e.g., atom) and categories (e.g., mammal) referenced in K-12 (NGSS Lead States, 2013) and college (College Board, 2009) science standards (Gampp et al., 2022). These entities and categories were then used to populate SW (Table 3.1), selecting a single entity for each power of ten from 10^{-10} m to 10^{9} m. A secondary entity (e.g., diameter of the Moon) was outlined on the floor of some worlds to increase student familiarity, especially for uncommonly referenced space objects (e.g., the dwarf planet Haumea). Within the environment, students can view and interact with representative entities at up to four orders of magnitude (e.g., acorn, American Robin, human, a Right Whale). SW underwent a principled development process, from mapping its features to the framework of conceptions of size and scale (Table 2.2) to multiple rounds of usability testing with experts, determined by the participants' experience in user interface design or graduate education with relevant experience in human factors (Wu et al., 2022). A different lesson plan focusing on energy production was implemented at the same participating middle school (Estrada et al., 2023), resulting in learning gains at a descriptive level as measured by the ASSC (Delgado et al., 2023), increased reform-oriented nature of instruction (Mathenge et al., 2023), and heightened student interest and motivation (Harper-Gampp et al., 2023c).

Table 3.1 Representative and secondary entities included in Scale Worlds VR.

Size	Representative Entity	Secondary Entity
10^{-10} m	Water Molecule	
10^{-9} m	Insulin Molecule	
10^{-8} m	Antibody	
10^{-7} m	COVID-19 Virus	
10^{-6} m	Mitochondrion	
10^{-5} m	White Blood Cell	
10^{-4} m	Paramecium	Thickness of a Human Hair
10^{-3} m	Ant	
10^{-2} m	Acorn	
10^{-1} m	Squirrel	
10^{0} m	Average Adult Human	
10^{1} m	Right Whale	
10^{2} m	International Space Station	American Football Field
10^{3} m	Brooklyn Bridge	
10^{4} m	Eros	State of Texas
10^{5} m	Haumea	Moon
10^{6} m	Agamemnon	
10^{7} m	Earth	
10^{8} m	Jupiter	
10^{9} m	Sun	

Delgado et al. (2023), in the same context as the dissertation, shows promise for SW to increase knowledge of size and scale among the target population. However, this earlier study utilized an adapted version of the ASSC, limiting the scope of size and scale reasoning assessed. Due to unforeseen circumstances, implementation was postponed to the last week of school following end-of-grade exams, which presented multiple challenges. For example, a change in location, shortened implementation timeline, and inconsistent participation across implementation days (Estrada et al., 2023). In addition, gains did not reach statistical significance (Delgado et al., 2023). For this dissertation study, I collaborated with the SW project on the deployment of SW.

Instructional Unit Exploring Human Impacts on Earth Systems

The partnership teacher and I iteratively developed a nine-day instructional unit (Table 3.2) investigating the socio-scientific issue (SSI) regarding the relationship between human systems and cropland, guided by the perspectives on strands or types of knowledge and their integration framework (Linn & Eylon, 2011; NGSS Lead States, 2013; NRC, 2012) and an emergent framework for SSI-based education (Sadler, 2011). SSI-based instruction should address issues of local concern (Sadler, 2011), and cropland loss is of particular interest because North Carolina is projected to lose 1.2 million acres of agricultural land (11.6%) by 2040 due to urbanization, one of the highest in the country (Xie et al., 2023). Blue Lake Middle School's county alone is projected to lose more than 25% (Xie et al., 2023). Therefore, we developed a unit that required students to take a multiscaled approach to some of the interactions between human systems and cropland. The unit consisted of nine 90-minute instructional periods with two days devoted to activity 1, one day devoted to activity 2, three days devoted to activity 3, and three days devoted to activity 4. A description of each instructional activity follows.

Activity 1: How has cropland changed near our school?

A recent implementation of SW-HMD at a local middle school demonstrated that students require support using the HMD equipment and navigating the virtual world (Harper-Gampp et al., 2023c). Meyer et al. (2019) demonstrated that the incorporation of pre-training had a positive impact on knowledge, transfer, and self-efficacy following the intervention. Therefore, the first instructional day incorporated time to ensure students are able to utilize the HMD equipment and navigate virtual environments. Given that North Carolina is projected to have one of the largest losses in cropland due to urbanization by 2040 in the country (Xie et al., 2023), students explored the agricultural industry within their state and investigated its impact on the local economy with Wander (Parkline Interactive, 2019), an application that allows users to teleport to anywhere in the world and observe changes over time.

Activity 2: What is the size of common science objects and our community?

Forming an integrated understanding of Earth systems is vital to improving life for all people (NASA, 2022). However, Hmelo et al. (2000) demonstrated that students' have difficulty recognizing that human systems operate across multiple scales, primarily focusing on visible aspects of a system (Mohan et al., 2009; Hmelo-Silver et al., 2007). Given that the development of temporal collective landmarks meaningfully scaffolds students' learning of scientific phenomena (Delgado, 2013a), I hypothesized that the establishment of spatial landmarks would support students' navigation of multiscale systems. Therefore, students utilized SW-HMD to construct a reference scale for entities from 10^{-10} m to 10^{9} m. I hypothesized that the embodied interactions with objects across scales would support students' conceptions of scale, which preliminary results have demonstrated (Delgado et al., 2023). Students completed tasks guided by the FS2C such as ordering objects by size, determining their absolute size, and comparing objects based on size (Magaña et al., 2012a). Collective landmarks at 10^{-10} m (atoms), 10^{-5} m (cells), 10^{0} m (human), 10^{5} m (width of North Carolina), 10^{2} m (length of their school), 10^{4} m (length of Wake County), 10^{9} m (Sun) were sought based on their salience in science curriculum and symmetry.

Activity 3: What does cropland need to thrive?

Cropland requires water to survive because of its vital role in photosynthesis, plant structure, and nutrient absorption (Duke Gardens, 2023). Globally, irrigation accounts for 70% of freshwater consumption (Integrate, 2018). However, Assaraf and Orion (2005) have demonstrated students' difficulty understanding the complex system and cyclic nature of water, and their inability to extend the system borders to include hidden dimensions (e.g., groundwater). Therefore, students explored three scale stories regarding water's connection to cropland: water for growth and development, water availability, and water treatment techniques. Scale stories establish explicit contents across dramatic changes in scale, such as from a crop to the water molecules supporting its growth and development, leveraging the power of narrative to assist students in establishing connections across scales (Garcia-Pelegrin et al., 2021). Students discussed these topics through whole-class instruction and participated in independent guided explorations. During instruction, student pairs intermittently utilized SW-HMD to elicit embodied experiences in connection with each scale story. For example, as we discussed the use of water for growth and development students started at the scale of a water molecule then progressed to a mitochondrion to draw connections between the scale of cellular respiration as a chemical process that is happening within a microscopic organelle. Students then grew to the size of a cell and a human to draw connections between microscopic entities' ability to support multicellular organisms. However, the collaborating teacher and I were faced with challenges given the varied support required. Therefore, a self-paced guided exploration was developed for students to explore water availability and water treatment techniques. For each scale story, students annotated their previously created reference scale to support the development of the previously identified collective landmarks.

Activity 4: How do humans impact cropland's ability to thrive?

A local concern is the health of the state's and county's rivers, which play a vital role in distributing fresh water. Therefore, students investigated the health and quality of their local river. Using their previously developed and annotated scales, students' constructed arguments for or against the development of local housing, industries, or farmland within the area. They were challenged with constructing a multiscale analysis of the impact of each of these changes on their local lake and rivers' health and quality. Students' constructed arguments were then disseminated to school administrators, who selected the highest quality argument.

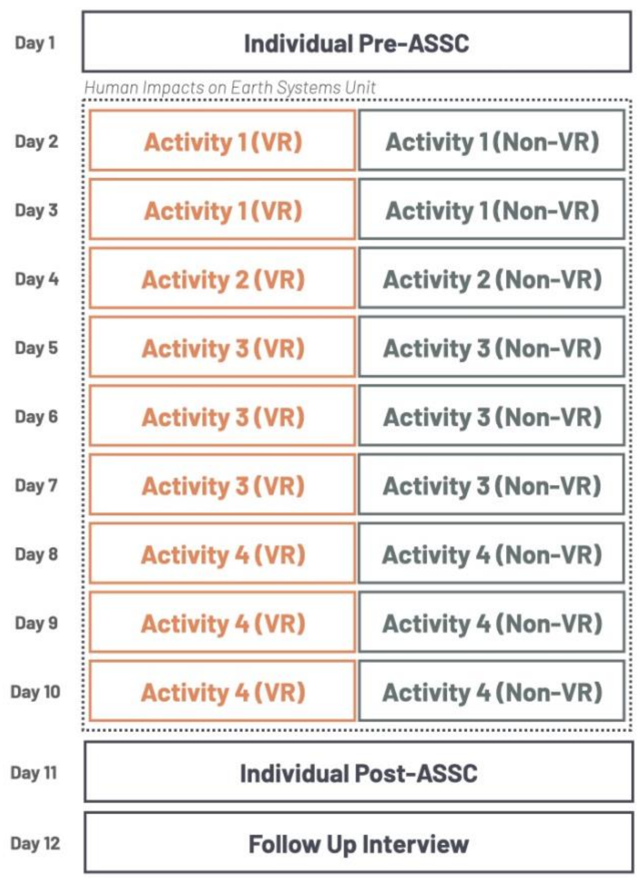
Table 3.2 Human impacts on Earth systems learning goals.

Sequence	**Learning Goal** (*Students will…)*
Activity 1: How has cropland changed near our school? (two days)	▪ Practice putting on the headset ▪ Create virtual boundaries with guardian ▪ Successfully cast to an additional device ▪ Use Wonder to scale from a global perspective to a students' local middle school and investigate changes over time ▪ Analyze historical aerial photographs of the area that illustrate urbanization in the local area ▪ Construct an explanation for how urbanization has impacted cropland
Activity 2: What is the size of common science objects and our community? (one day)	▪ Establish accurate individual landmarks (i.e., molecule, cell, human, North Carolina, & the Sun) using Scale Worlds VR ▪ Identify the size of common objects referenced in national science standards ▪ Compare the relative size and a scale depicting common objects reference in national science standards
Activity 3: What does cropland need to thrive? (three days)	▪ Describe crops use of water for growth and development ▪ Analyze the use of water to support crop growth and development across three different perspectives: water availability (10^7 m to 10^4 m), water for growth and development (10^{-10} m to 10^0 m), and water treatment (10^{-10} m to 10^2 m).
Activity 4: How do humans impact cropland's ability to thrive? (three days)	▪ Describe the vital role rivers play in crop growth and development ▪ Analyze human impacts on rivers within the state ▪ Construct an argument for methods that support a balanced relationship among rivers, cropland, and society ▪ Disseminate arguments to school administrators for scoring

Study 1: Investigating Middle School Students' Scale Reasoning Through Immersive Virtual Reality-Supported Science Instruction

An explanatory mixed methods design (Figure 3.4) was utilized to explore changes in scale reasoning through the incorporation of SW-HMD, an immersive virtual environment that allows students to explore the extremes of scale. It was hypothesized that the direct and embodied experiences facilitated by SW-HMD would support students' scale reasoning, which research has demonstrated are difficult for students (e.g., Tretter et al., 2006b). Quantitative data was collected followed by qualitative data (QUAN→qual) to enrich experimental results as recommended by Creswell & Plano Clark (2019) given that research on VR interventions at the middle school level are still emerging (Luo et al., 2021; Matovu et al., 2023) and a prior study on SW-HMD demonstrated a positive impact on scale reasoning but gains did not reach statistical significance (Delgado et al., 2023). At baseline and conclusion, students completed the Assessment of Size and Scale Cognition (ASSC). Following the intervention, a subset of students were interviewed to explain quantitatively observed changes in scale reasoning after using SW-HMD. The quantitative outcome results were combined with the qualitative results to bring greater insight into the interplay between the affordances of SW-HMD and student cognitive outcomes.

Figure 3.2 Illustration of study one's research design.



Research Questions

As recommended by Creswell and Plano Clark (2018), I identify a quantitative, qualitative, and mixed methods research question to guide this study.

- How does the size and scale reasoning of middle school students' change after VR-supported science instruction? (QUAN)
- What role does Scale Worlds VR play in students' development of size and scale reasoning? (qual)
- How do middle school students' interviews help to explain quantitatively observed changes in scale reasoning after VR-supported science instruction? (QUAN→qual)

Data Collection

Prior to the intervention, students completed the Assessment of Size and Scale Cognition (ASSC; Harper-Gampp et al., 2023a, 2023b,2023d), a validated scale cognition assessment aligned with three of the five fundamental ways of thinking about size and scale proposed by Magaña et al. (2012a). Students then participated in a nine-day instructional unit exploring human impacts on Earth systems. For one class, VR was incorporated throughout the unit as an instructional tool, purposefully integrated within the first three activities and available for use for the last one. Time was devoted on the first instructional day to increase students' familiarity with the VR equipment and navigating VR environments as recommended by Meyer et al. (2019) to mitigate high cognitive demand observed in our earlier implementation (Harper-Gampp et al., 2023c). The other class utilized Universcale (Nikon, 2009), a previously identified tool for supporting students' size and scale reasoning (Magaña, 2014), and the web-based version of SW.

At the conclusion of the unit, students completed the ASSC, a feedback survey (Appendix A), and interested students were invited to participate in a semi-structured interview (Appendix B) regarding their experience with SW-HMD and the alternative instructional tools (i.e., Unvierscale, SW-Web). Interview questions were created to elicit details regarding individual perspectives about their experience (e.g., describe your use of the various measurement tools in SW to learn more about the size of a water molecule). Open-ended questions regarding the experience were included to further understand the impact of SW on the development of size and scale reasoning. A subset of students (n = 11) were interviewed to elicit Individual perspectives as recommended by Fofana et al. (2020).

Data Analysis

Scoring Assessment. Scores were determined for each section of the ASSC (ordering, absolute, and relative) and then summed for a composite score. For the ordering task, each entity was coded as correct (1 point) if the entity directly above and below were correct, or incorrect (0 points). For example, if a student ordered the school bus (10^1 m) directly above the human (10^0 m) and the squirrel (10^{-1} m) directly below the human, human was scored as correct (1 point). However, if they ordered the bridge (10^3 m) above the school bus, school bus was scored as incorrect (0 points). For each absolute and relative task, responses were recorded as correct (2 points) if the correct entity/size was selected, partial (1 point) if the entity/size was off by one order of magnitude, or incorrect (0 points).

Pre-analyses. Of the 36 participating students (n_{VR} = 20, n_{NVR} = 16), 34 students completed the pre-ASSC (n_{VR} = 20, n_{NVR}= 14), 34 students took the post-ASSC (n_{VR} = 22 , n_{NVR} = 12), and 26 matched pairs were obtained (n_{VR} = 18 , n_{NVR} = 8). One non-VR student's paired responses were excluded based on evidence of outside resource usage resulting in 8 matched pairs. Time spent completing each section of the assessment was used to identify quality responses. If a student took 20 seconds or less to complete a section their responses for that section were excluded, leading to 24 (n_{VR} = 16 , n_{NVR} = 8) matched pairs for ordering, 25 (n_{VR} = 17 , n_{NVR} = 8) matched pairs for absolute, 18 (n_{VR} = 13 , n_{NVR} = 5) matched pairs for relative, and 15 complete matched pairs (n_{VR} = 10 , n_{NVR} = 5).

Comparison. The pre- and post-ASSC composite score and within section scores for the VR and non-VR condition were statistically compared with a mixed ANOVA. An ANOVA test requires an approximately normal distribution. Shapiro-Wilk's tests indicated that each sample overall and within each section was normally distributed ($p > 0.05$). Group differences prior to instruction were assessed through independent sample t-tests, which indicated no statistically significant differences in mean pre-test scores between groups. This indicates no significant difference in size and scale reasoning among groups prior to instruction. Mean scores were compared within groups by conducting a paired samples t-test. Cohen's *d* indices were then calculated to assess effect sizes. Collective landmarks, alternative ideas, and instances of gaps in knowledge for both absolute and relative size values were calculated following the procedure outlined by Delgado (2013a), as previously described. To situate results among previous findings, the average rank for ordering, absolute, and relative were determined following procedures outlined by Tretter et al. (2006a).

Interviews. Interviews were analyzed using the constant comparative method (Corbin & Strauss, 2008) in which the empirical data is reviewed in search of emergent themes. Initially, the data was coded utilizing a set of a priori VR affordances (Steffen et al., 2019; Lee, 2004) followed by a round of open-coding to identify emergent codes. Results were consolidated into a codebook (Table 3.3) and example responses for each observed code were added to increase clarity.

To ensure inter-rater reliability, an individual with a doctorate in science education was recruited to code a representative sample (10%) using the developed codebook. The level of agreement between the two raters yielded a Cohen's Kappa of 0.75, which is considered substantial agreement (Cohen, 1960; Landis & Koch, 1977).

Table 3.3 Codebook of a priori affordances of virtual reality.

Code	Definition	Example(s)
Diminish Negative Aspects	"Virtual representations of activities that normally have large negative effects but those effects are artificially diminished (Steffen et al., 2019, p. 696)"	Not observed
Enhance Positive Aspects	"Virtual representations are used to present information in a method that improve upon methods available in physical reality (Steffen et al., 2019, p. 696)"	*(1) Helped me learn, (2) you can actually see it in the VR, (3) use VR to see the surroundings of the story*
Recreate Existing Aspects	"Virtual representations are used to replicate some object or experience in the real world, not because it is better represented virtually, but because it is more practical to do so (Steffen et al., 2019, p. 696)"	*Not observed*
Create Aspects	"Virtual representations are used because reasonable physical counterparts simply do not exist (Steffen et al., 2019, p. 696)"	*(1) you can grow like as big or as small as you can; (2) we could see how bigs things were from our point of view*
Presence	A sense of "realness" – these experiences might be para-authentic (e.g., exploring a power plant) or artificial (e.g., experiencing a historic battle) (Lee, 2004, p. 37)."	*(1) engulfed; (2) actually in it*

Study 2: Exploring Human Impacts on Earth Systems Across Scale Worlds in Immersive Virtual Reality: An Investigation of Students' Integrated Understanding

A quasi-experimental design (Figure 3.5) was utilized to explore changes in students' integration across CCC and multiple DCIs. It was hypothesized that the direct and embodied experiences facilitated by SW-HMD would support students' knowledge integration, which research has demonstrated is difficult for students (e.g., Assaraf and Orion, 2005). At baseline and conclusion, students completed an SPQ and DCI knowledge integration assessment (Table 3.4). The assessment consisted of ten free response items aligned to instruction and piloted among a small group of the sample population. Survey responses were first coded for the presence of a DCI and SPQ. Following a constant comparative method (Corbin & Strauss, 2008), connections across integrated responses, those coded for both a DCI and SPQ, were inductively analyzed to extract patterns, categories, and themes. Integrated responses were then compared across groups. A chi-squared test of independence was utilized to investigate the frequency of integrated responses across groups. Following the intervention, a subset of students were interviewed to gain a deeper understanding of students' integration across CCC and multiple DCI after using SW-HMD.

Figure 3.3 Illustration of study two's research design.

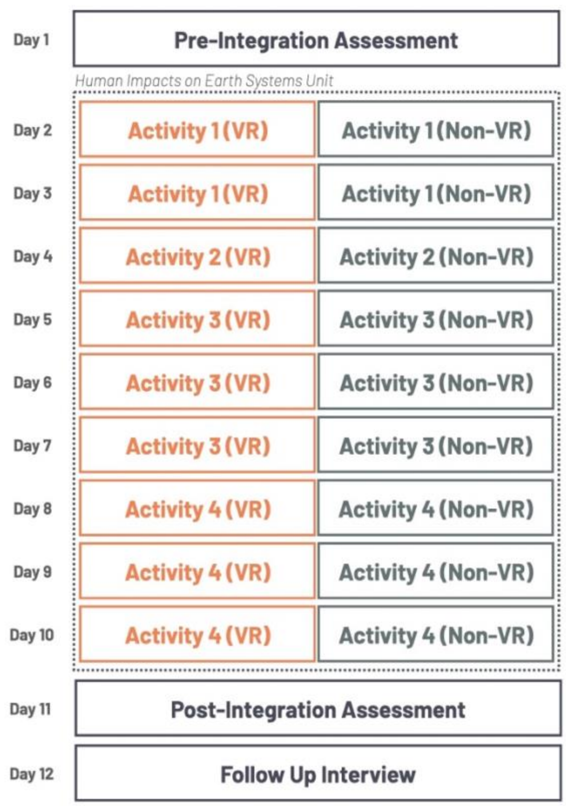


Table 3.4 SPQ and DCI integration assessment items.

Activity	DCI	Item
Activity 1: How has cropland changed near our school?	**ESS3.C** Human Impacts on Earth Systems	How do human activities impact cropland? How does city growth impact cropland on a cellular and global scale?
Activity 3: What does cropland need to thrive?	**LS1.C** Organization for Matter and Energy Flow Organisms	How does water help plants grow and survive?
	ESS2.C The Roles of Water in Earth's Surface Processes	What are some different ways a crop gets water? How are water molecules in the Earth's oceans connected to the water molecules in a plant cell?
Activity 4: How do humans impact cropland's ability to thrive?	**ESS3.C** Human Impacts on Earth Systems	What are some things I might find in polluted water? What are some things I might find in polluted water that are different in size? How do humans remove these different pollutants from polluted water? A farmer has chosen to use fertilizer to support their crops. How does this choice impact the quality of water? A farmer has chosen to use fertilizer to support their crops. How can this impact different organisms, ecosystems, local communities, and global communities?

Research Questions

The specific questions that guided this study are as follows:

- How do middle school students integrate a crosscutting concept and a disciplinary core idea?
- How does immersive virtual reality-supported instruction influence middle school students' integrated understanding of a crosscutting concept and a disciplinary core idea?

Data Collection

Prior to the intervention, students completed a two-dimensional knowledge integration assessment (Table 3.4), consisting of ten free response items aligned to instruction that were developed to elicit the integration of SPQ and multiple DCIs implicitly and explicitly. For the DCI concerning human impacts on the biosphere, students were asked to describe an impact in general (item 1) followed by item 2, which explicitly asked students to consider impacts across scales (i.e., on a cellular and global scale). An initial draft of items was piloted among a small group of the sample population and modified for clarity. Students then participated in the nine-day instructional unit exploring human impacts on Earth systems previously described. At the conclusion of the unit, students completed the assessment and interested students were invited to participate in a semi-structured interview (Appendix A) regarding their experience with SW. Interview questions elicited details regarding individual perspectives about their experience (e.g., what part of the activities do you feel helped you to understand these concepts?). A subset of items (1, 3, & 5) from the assessment were included in the interview protocol to further understand the impact of SW on the development of an integrated understanding of a CCC and DCI. A subset of students (n = 11) were interviewed to elicit individual perspectives as recommended by Fofana et al. (2020).

Data Analysis

Student responses were analyzed in aggregate using the constant comparative method (Corbin & Strauss, 2008) in which the empirical data is reviewed in search of emergent themes. Initially, each response was coded utilizing a set of a priori disciplinary core ideas (DCIs; Table 3.5) identified in the Framework for K-12 Science Education (NRC, 2012) with which the assessment items were developed (Table 3.4). A set of rules applied during the coding process with detailed rationales are included in Appendix C. This was followed by a round of open coding to identify additional DCIs.

Each response was then coded for the CCC of scale (S) and proportion and quantity (PQ) as operationalized by Laverty et al., 2016 (Table 3.6). Scale ranges had to be established to determine if comparisons across scales had been made, otherwise any comparison of even the most minute scale would be considered (e.g., a large cup and a small cup). Tretter et al. (2006b) had previously observed a sharp decline in accuracy among learners of varied expertise and age. Therefore, anything smaller than the visible limit (<10^{-3} m) was established as a range. Large scale, referred to as big scale based on conceptual boundaries identified by Tretter et al. (2006a), was defined as anything larger than a kilometer because previous findings indicated that people's conception of size tend to be relative to their own size (Tretter et al., 2006a). This led to ranges that were consistent with boundaries defined on the Scale Anchoring Objects (SAO) assessment and supported by subsequent findings by Trate et al. (2019). Therefore, student responses were coded for scale if comparisons between the identified ranges were made. For example, a comparison between a cell and a human was coded for scale while a comparison between a fruit fly and a bumblebee was not coded for scale. Example responses for each observed code were added to increase clarity (Table 3.6). A set of rules applied during the coding process with detailed rationales are included in Appendix D.

An integrated response was defined as those including both a DCI and the CCC of SPQ. Therefore, responses coded for a DCI and either S, PQ or S and PQ were isolated and reviewed in search of emergent themes. To ensure inter-rater reliability, an individual with a doctorate in science education was recruited to code a representative sample (10%) using the developed codebook for DCIs and SPQ. The level of agreement between the two raters yielded a Cohen's Kappa of 0.77 for DCIs and 0.71 for SPQ, which are both considered substantial agreement (Cohen, 1960; Landis & Koch, 1977).

Table 3.5 Codebook of a priori disciplinary core ideas including definitions and examples. Additional rules and detailed rationales are included in Appendix C.

Code	Definition	Example(s)
LS1.C Organization for Matter and Energy Flow in Organisms	"Plants need water and light to live and grow…Energy needed for life is ultimately derived from the sun through photosynthesis. Plants, algae (including phytoplankton), and other energy fixing microorganisms use sunlight, water, and carbon dioxide to facilitate photosynthesis (NRC, 2012, p. 147)"	*(1) they need water and food to stay alive; (2) water long with soil and sunlight help the plant grow and live*
ESS2.C The Roles of Water in Earth's Surface Processes	"Water is found almost everywhere on Earth, from high in the atmosphere (as water vapor and ice crystals) to low in the atmosphere (precipitation, droplets in clouds) to mountain snowcaps and glaciers (solid) to running liquid water on the land, ocean, and underground. Energy from the sun and the force of gravity drive the continual cycling of water among these reservoirs (NRC, 2012, p. 184)"	*(1) because the water is all over the world and water can be on anything; (2) Plants can get water from rainfall.*
ESS3.C Humans Impacts on Earth Systems	"Human activities in agriculture, industry, and everyday life have had a major impact on the land, rivers, ocean, and air…and because all of Earth's subsystems are interconnected, changes in one system can produce unforeseen changes in others (NRC, 2012, p. 195)." This includes the "distribution of Earth's water through the modification of streams, lakes, and groundwater (NRC, 2012, p. 194)." "Some negative effects of human activities are reversible with informed and responsible management. For example, communities are doing many things to help protect Earth's resources and environments. They are treating sewage, reducing the amount of materials they use, and reusing and recycling materials (NRS, 2012, p. 195)."	*(1) city growth impacts cropland because we are taking away soil and the population is growing; (2) it's too many city's putting gas in the air and it kills plants*

Table 3.6 Codebook for scale, proportion, and quantity including definitions and examples. Additional rules and detailed rationales are included in Appendix D.

Code	Definition	Example(s)
Scale	(1) the comparison of objects, processes, or properties across size, time, or energy scales, or to dimensions of familiar objects, timescale, or energies or (2) the identification of non-negligible/relevant interactions at more than one scale (Laverty et al., 2016, p. 8)	a) depending how you grow your plants (*human scale*) [water] can help give plants nutrients (*microscopic scale*) b) contaminating (*human scale*) the soil so crops can't grow, polluting (*human scale*) the land and water, and urbanizing (*big scale*) areas [*Actions were initially assumed to be human scale based on the wording of the item (e.g., polluting)*] c) dirt, trash (*human scale*), bacteria (*microscopic scale*), decomposed material
Proportion and Quantity	(1) identifying a quantitative response of one variable to changes in another or (2) the relationship between two or more quantitative variables (Laverty et al., 2016, p. 8). "Grasp the relationship among different types of quantities" which is "different than a ratio of numbers describing fractions of a pie (NRC, 2012, p. 90)"	a) city growth can impact it on a cellular scale by making more electricity and it can impact it on a global scale by taking away places to plant (*more city leads to less places to plant*) b) the fertilizer can become runoff waste if there's too much water that the plant can handle. (*too much water leads to increase in runoff*)

CHAPTER 4: Results

Study 1: Investigating Middle School Students' Scale Reasoning Through Immersive Virtual Reality-Supported Science Instruction

Research Question 1A

How do middle school students' scale reasoning change after VR-supported science instruction? (QUAN)

Assessment of Size and Scale Reasoning (ASSC). The ASSC, a previously validated instrument (Harper-Gampp et al., 2024a, 2024b, 2024d), was administered at baseline and conclusion to explore how students' size and scale reasoning changed following instruction. There was a significant increase in students' composite score on the ASSC following VR enriched instruction (M = 30.70, SD = 10.41) compared to students' initial composite score (M = 24.20, SD = 9.81), $t(9)$ = 4.52, $p < 0.005$ with a large effect size (d = 0.6). Similarly, a significant increase in students' composite score was observed following non-VR instruction (M = 27.80, SD = 11.12) compared to students' initial composite score (M = 20.80, SD = 9.52), $t(4)$ = 4.43, $p < 0.05$ with a medium effect size (d = 0.68). On average, students participating in VR enriched instruction improved by 6.5 points while non-VR students' performance improved by 7.0 points. Students participating in VR enriched instruction significantly improved across all three fundamental ways of thinking about size and scale assessed on the ASSC (Table 4.1): ordering ($p < 0.05$), absolute ($p < 0.05$), and relative ($p < 0.005$). While those participating in non-VR instruction descriptively improved across ordering (+3.4 pts), absolute (+3.1 pts), and relative (+2.8 pts), results were not statistically significant (Table 4.2). However, Cohen's d indices for each section ranged from 0.6 to 1.0, which Cohen (1988) describes as a medium to large effect.

Table 4.1 Comparing VR students' size and scale reasoning performance as assessed by the ASSC prior to and following instruction.

			Pre-Test		Post-Test	
	n	*d*	*M*	*SD*	*M*	*SD*
Composite Score (Out of 86)	10	0.64	24.20	9.81	30.70**	10.41
Ordering (Out of 10)	13	0.38	4.06	4.54	5.25*	2.72
Absolute (Out of 40)	17	0.38	10.53	6.41	13.18*	7.52
Relative (Out of 36)	16	0.60	8.54	3.47	11.15**	4.14

*$p < 0.05$ **$p < 0.005$

Table 4.2 Comparing non-VR students' size and scale reasoning performance as assessed by the ASSC prior to and following instruction.

			Pre-Test		Post-Test	
	n	*d*	*M*	*SD*	*M*	*SD*
Composite Score (Out of 86)	5	0.63	20.80	9.52	27.80*	11.12
Ordering (Out of 10)	8	1.05	2.50	2.83	5.88	3.56
Absolute (Out of 40)	8	0.56	7.13	5.72	10.25	5.39
Relative (Out of 36)	5	0.58	8.80	4.84	11.60	4.83

*$p < 0.05$ **$p < 0.005$

Table 4.3 Comparing VR and non-VR students' size and scale reasoning performance as assessed by the ASSC following instruction.

			VR Post-Test		Non-VR Post-Test	
	d	*Δ M*	*M*	*SD*	*M*	*SD*
Composite Score (Out of 86)	0.27	2.90	30.70	10.41	27.80	11.12
Ordering (Out of 10)	-0.20	-0.63	5.25	2.72	5.88	3.56
Absolute (Out of 40)	0.45	2.93	13.18	7.52	10.25	5.39
Relative (Out of 36)	-0.10	-0.45	11.15	4.14	11.60	4.83

*p < 0.05 **p < 0.005

Descriptively, students participating in VR enriched instruction performed better than non-VR participating students overall (Table 4.3). However, a mixed ANOVA revealed that these differences were not statistically significant, $F(3,26) = 1.32$, $p = 0.29$. Similarly, VR students performed descriptively better in one of the three sections (absolute). However, a mixed ANOVA revealed that these differences were not statistically significant, $F(3,46) = 1.60$, $p = 0.20$. A small effect size was observed for both the composite score and absolute task items. Students assigned to the non-VR condition showed better performance on the ordering and relative task items as compared to the students assigned to the VR condition, but differences were not statistically significant with a small effect size ($d = 0.2$).

Ordering. Students' entity ranking was averaged for VR and non-VR participants prior to and following the intervention (Table 4.4). Prior to instruction, the VR participants' longest correctly ordered set contained three entities (Golden Gate Bridge, Titanic, & school bus). No correctly ordered set was identified for non-VR participants prior to instruction. Following instruction, the VR participants' longest correctly ordered set contained five entities (Mount Everest, Golden Gate Bridge, Titanic, school bus, human) and non-VR participants' longest correctly ordered set contained three entities (bumblebee, fruit fly, human egg cell). However, both groups' correctly ordered sets contained mostly human scale entities (10^3 m to 10^{-3} m), with VR students including one entity larger and non-VR including one entity smaller.

On the pre-test, VR participants ranked three entities smaller than the smallest-ranked visible object (fruit fly) and non-VR participants correctly ranked two entities smaller than the smallest-ranked visible object (fruit fly). Both groups improved following instruction, with VR participants correctly ranking all seven entities smaller than the smallest-ranked visible object (fruit fly) and non-VR participants correctly ranking four entities smaller than the smallest-ranked visible object (fruit fly). In addition, no entities too small to see were correctly ranked (i.e., if the entity falling above and below were one order-of-magnitude higher and lower, respectively) among either group.

Table 4.4 VR and non-VR students' average ranking for each entity prior to and following the instructional unit.

		VR		**non-VR**	
Entity	**Actual Rank**	*Pre-Rank*	*Post-Rank*	*Pre-Rank*	*Post-Rank*
Sun (Diameter)	**1**	1	1	4	2
Saturn (Diameter)	**2**	2	2	2	1
Venus (Diameter)	**3**	5	3	11	4
Mercury (Diameter)	**4**	3	4	7	6
Panama Canal (Length)	**5**	9	10	13	9
Mount Everest (Height)	**6**	4	5	1	8
Golden Gate Bridge (Length)	**7**	6	6	3	5
Titanic (Length)	**8**	7	7	5	3
School Bus (Length)	**9**	8	8	8	7
Human (Height)	**10**	10	9	6	10
Squirrel (Height)	**11**	12	11	9	16
Bumblebee (Length)	**12**	11	12	14	11
Fruit Fly (Length)	**13**	17	13	18	12
Human Egg Cell (Diameter)	**14**	14	14	20	13
Skin Cell (Diameter)	**15**	18	17	17	17
Nucleus of a Cell (Diameter)	**16**	20	19	15	19
Influenza Virus (Diameter)	**17**	16	20	19	14
Ribosome (Length)	**18**	13	15	12	15
Glucose Molecule (Length)	**19**	15	18	16	18
Carbon Dioxide Molecule (Length)	**20**	19	16	10	20

Absolute. Following Delgado's (2013a) procedure, collective landmarks for absolute size were calculated prior to and following instruction for both VR (Table 4.5 - 4.6) and non-VR students (Table 4.7 - 4.8). The Sun was the only entity that met the criteria for both accuracy and precision, establishing it as a collective landmark, among VR students following instruction. No entities met the criteria for both accuracy and precision prior to or following instruction for non-VR students.

Prior to instruction, 18 entities for VR students and 11 entities for non-VR students met the criteria for accuracy but not precision. VR students also demonstrated a more accurate conception of multiple entities (e.g., skin cell (35%), Sun (30%), Mercury (15%), Golden Gate Bridge (10%)) with no more than 5% of non-VR students estimating the correct size for any entity. Following instruction, only 14 entities for VR students and 16 entities for non-VR students met the criteria for accuracy but not precision. VR students' accuracy declined for several small entities, including bumblebee, skin cell, influenza virus, ribosome, and glucose. While non-VR students demonstrated improved precision among multiple entities close in proximity to human scale (e.g., squirrel).

An entity meeting the criteria for precision but not accuracy suggests an alternative idea. No entities among VR or non-VR students prior to or following instruction met this criterion. However, following instruction, 30% of VR and non-VR students estimated the absolute size of the fruit fly to be one order-of-magnitude larger, possibly considering the size of a common housefly or horsefly. In addition, 35% of non-VR students estimated the absolute size of Venus to be three orders-of-magnitude smaller compared to only 10% of VR students. This is surprising given Venus was not included in SW but was part of Universcale. Similarly, 30% of non-VR students estimated the absolute size of Mount Everest to be two orders-of-magnitude smaller compared to less than 5% of VR students, which is also included in Universcale and not SW. This might be attributed to the way absolute size is reported in SW versus Universcale. In SW, the absolute size of each entity is presented in scientific and standard notation (e.g., 1.3×10^3 m and 1.3 km) while only standard notation is reported in Universcale. The Panama Canal was the only entity prior to and following instruction among both groups that neither met accuracy or precision criteria. This suggests that there is lack of knowledge among all students regarding the size of this entity or lack of familiarity. Absolute size values for both VR and non-VR students prior to and following instruction from the 5th to 95th percentiles can be found in Appendix E.

Table 4.5 Pretest absolute size values for VR participating students from the 25th through 75th percentiles. Correct values for each entity are colored blue with potential alternative ideas colored grey.

Entity	25	30	35	40	45	50	55	60	65	70	75	M
Sun (Diameter)	3	4	6	8	9	9	9	9	9	9	9	6.1
Saturn (Diameter)	3	3	4	5	7	8	8	9	9	9	10	5.9
Venus (Diameter)	2	2	3	4	4	4	4	5	7	8	8	4.4
Mercury (Diameter)	4	5	6	6	6	6	7	7	7	8	9	5.9
Panama Canal (Length)	2	2	3	3	3	3	3	4	4	4	4	2.2
Mount Everest (Height)	4	5	6	6	6	6	7	7	8	9	10	6.2
Golden Gate Bridge (Length)	1	1	1	1	2	2	3	3	3	4	5	2.7
Titanic (Length)	2	3	3	3	3	5	5	5	5	6	6	4.0
School Bus (Length)	0	0	1	1	1	1	1	1	1	1	2	0.6
Human (Height)	-3	-2	-1	-1	-1	-1	0	0	0	0	0	-1.0
Squirrel (Height)	-5	-4	-3	-3	-3	-2	-2	-1	-1	-1	-1	-2.8
Bumblebee (Length)	-9	-7	-6	-5	-3	-2	-2	-2	-2	-2	-1	-4.1
Fruit Fly (Length)	-8	-6	-5	-5	-5	-4	-3	-3	-3	-2	-2	-4.9
Human Egg Cell (Diameter)	-10	-9	-8	-7	-7	-5	-5	-5	-5	-4	-4	-6.3
Skin Cell (Diameter)	-6	-5	-5	-5	-5	-5	-5	-5	-5	-4	-3	-3.9
Nucleus of a Cell (Diameter)	-10	-8	-7	-6	-6	-6	-5	-4	-4	-4	-4	-5.6
Influenza Virus (Diameter)	-8	-8	-7	-7	-6	-5	-5	-5	-5	-4	-3	-5.2
Ribosome (Length)	-9	-9	-9	-9	-8	-7	-7	-7	-7	-6	-5	-5.9
Glucose Molecule (Length)	-10	-10	-9	-9	-8	-8	-6	-6	-6	-5	-3	-5.5
Carbon Dioxide Molecule (Length)	-8	-7	-7	-7	-7	-6	-6	-5	-5	-4	-4	-4.3

Table 4.6 Posttest absolute size values for VR participating students from the 25th through 75th percentiles. Correct values for each entity are colored blue with potential alternative ideas colored grey.

Entity	25	30	35	40	45	50	55	60	65	70	75	M
Sun (Diameter)	9	9	9	9	9	9	9	9	9	9	9	9.2
Saturn (Diameter)	4	5	5	5	6	6	6	7	7	7	7	4.5
Venus (Diameter)	4	4	4	5	6	7	7	7	7	7	8	4.9
Mercury (Diameter)	3	4	5	5	5	5	6	6	6	6	7	5.0
Panama Canal (Length)	-4	-1	1	1	2	4	4	4	4	4	4	0.5
Mount Everest (Height)	3	3	4	4	4	4	4	4	4	5	5	3.5
Golden Gate Bridge (Length)	3	3	3	3	3	4	4	4	4	5	7	4.8
Titanic (Length)	1	1	1	1	2	2	3	3	3	3	5	2.9
School Bus (Length)	0	0	1	1	1	1	1	1	1	1	1	-0.4
Human (Height)	-3	-2	-1	-1	-1	0	0	0	0	0	0	-1.6
Squirrel (Height)	-3	-3	-2	-2	-2	-1	-1	-1	-1	-1	-1	-1.5
Bumblebee (Length)	-6	-5	-4	-3	-3	-3	-1	-1	-1	-1	-1	-2.8
Fruit Fly (Length)	-4	-3	-3	-3	-2	-2	-2	-2	-2	-2	-2	-2.9
Human Egg Cell (Diameter)	-7	-5	-5	-5	-5	-5	-4	-4	-4	-4	-3	-4.3
Skin Cell (Diameter)	-7	-6	-6	-6	-6	-4	-3	-3	-3	-3	-2	-3.7
Nucleus of a Cell (Diameter)	-9	-8	-8	-8	-8	-7	-6	-5	-5	-4	-3	-4.8
Influenza Virus (Diameter)	-9	-9	-8	-7	-7	-7	-7	-7	-7	-6	-6	-6.2
Ribosome (Length)	-9	-8	-8	-8	-8	-8	-6	-5	-5	-5	-4	-5.8
Glucose Molecule (Length)	-8	-7	-7	-7	-7	-6	-6	-6	-6	-6	-4	-5.2
Carbon Dioxide Molecule (Length)	-8	-8	-7	-6	-6	-6	-6	-5	-5	-5	-4	-5.4

Table 4.7 Pretest absolute size values for non-VR participating students from the 25th through 75th percentiles. Correct values for each entity are colored blue with potential alternative ideas colored grey.

Entity	25	30	35	40	45	50	55	60	65	70	75	M
Sun (Diameter)	6	7	7	7	7	8	8	8	8	**9**	**9**	5.2
Saturn (Diameter)	5	7	**8**	10	10	11	11	11	11	11	11	7.8
Venus (Diameter)	2	2	2	2	2	3	3	3	4	5	5	1.7
Mercury (Diameter)	8	8	8	8	9	10	10	11	11	11	11	8.5
Panama Canal (Length)	-1	-1	0	1	2	2	3	3	3	4	4	1.0
Mount Everest (Height)	-5	-3	-2	0	0	1	1	1	2	**4**	5	0.3
Golden Gate Bridge (Length)	4	4	4	4	5	6	6	7	8	8	9	6.5
Titanic (Length)	-1	-1	0	1	**2**	**2**	3	3	3	3	3	1.2
School Bus (Length)	-6	-5	-5	-4	-3	-2	-1	0	0	0	0	-3.0
Human (Height)	-3	-3	-2	-2	-2	-2	-2	-2	-1	**-0**	1	-1.3
Squirrel (Height)	-6	-6	-5	-5	-4	-3	**-1**	-0	0	1	1	-2.2
Bumblebee (Length)	-9	-8	-8	-7	-7	-6	-6	-5	-5	-5	-5	-5.7
Fruit Fly (Length)	-9	-9	-9	-9	-9	-9	-8	-8	-7	-6	-5	-7.2
Human Egg Cell (Diameter)	-10	-10	-10	-10	-8	-6	**-4**	-2	-1	-1	0	-5.3
Skin Cell (Diameter)	-9	-9	-8	-8	-7	-7	-6	**-5**	-3	-1	2	-3.2
Nucleus of a Cell (Diameter)	**-6**	-5	-5	-4	-3	-2	-0	1	1	2	2	-2.7
Influenza Virus (Diameter)	-10	-8	-6	-4	-4	-4	-3	-3	-3	-3	-2	-4.7
Ribosome (Length)	-1	-1	-1	-1	-0	1	1	2	2	3	3	0.3
Glucose Molecule (Length)	-10	**-9**	-7	-6	-5	-5	-4	-3	-3	-2	-2	-5.2
Carbon Dioxide Molecule (Length)	-11	**-10**	-8	-7	-4	-1	2	5	5	5	5	-2.0

Table 4.8 Posttest absolute size values for non-VR participating students from the 25th through 75th percentiles. Correct values for each entity are colored blue with potential alternative ideas colored grey.

Entity	25	30	35	40	45	50	55	60	65	70	75	M
Sun (Diameter)	7	7	7	7	8	8	9	9	9	9	9	7.8
Saturn (Diameter)	5	7	8	9	9	9	9	9	9	10	10	7.3
Venus (Diameter)	3	3	4	4	4	4	4	4	4	5	5	4.0
Mercury (Diameter)	2	3	3	3	4	4	5	5	5	5	5	3.8
Panama Canal (Length)	-0	1	1	2	2	2	2	2	2	3	3	1.7
Mount Everest (Height)	1	2	2	2	2	2	2	2	3	4	4	2.8
Golden Gate Bridge (Length)	-1	1	2	4	4	4	4	4	5	6	7	2.8
Titanic (Length)	0	1	2	3	4	4	5	5	6	6	7	2.5
School Bus (Length)	0	1	1	1	1	1	1	1	1	2	2	1.5
Human (Height)	-2	-1	-1	0	0	0	0	0	0	0	0	-0.8
Squirrel (Height)	-5	-5	-5	-5	-4	-3	-2	-1	-1	-1	-1	-3.2
Bumblebee (Length)	-3	-3	-2	-2	-2	-2	-2	-2	-1	-1	0	-0.2
Fruit Fly (Length)	-5	-4	-3	-2	-2	-2	-2	-2	-2	-2	-1	-3.5
Human Egg Cell (Diameter)	-8	-8	-7	-7	-7	-6	-6	-5	-5	-4	-4	-5.2
Skin Cell (Diameter)	-5	-5	-4	-4	-4	-4	-3	-3	-3	-2	-2	-3.2
Nucleus of a Cell (Diameter)	-7	-7	-6	-5	-3	0	3	5	6	6	7	-0.7
Influenza Virus (Diameter)	-9	-8	-8	-7	-6	-5	-3	-2	-2	-1	-1	-3.7
Ribosome (Length)	-9	-9	-8	-7	-6	-5	-4	-3	-3	-2	-2	-4.5
Glucose Molecule (Length)	-10	-10	-9	-9	-8	-7	-5	-4	-2	-1	1	-4.3
Carbon Dioxide Molecule (Length)	-11	-11	-10	-10	-10	-10	-9	-9	-8	-7	-6	-7.5

Absolute rankings of entities were determined by ranking the mean size values for a given entity for VR (Table 4.9) and non-VR (Table 4.10) students prior to and following instruction. Prior to instruction, VR students correctly ranked all entities that were larger than human larger and entitles that were smaller than human smaller. However, following instruction, squirrel was ranked larger than a human. The standard deviation of the absolute ranks indicate that this is likely due to students increased variability in the height of a human with students indicating this size an order-of-magnitude smaller or more. Following instruction, VR students demonstrated improved precision for most entities with an average variability of

4.09 compared to an initial 4.47. VR students longest correctly ordered set contained two entities on the pre-test (Sun and Saturn; human and squirrel) and post-test (bumblebee and fruit fly; glucose and carbon dioxide).

Prior to instruction, non-VR students incorrectly ranked a ribosome larger than a human and a school bus smaller that a human. Following instruction, non-VR students incorrectly ranked a bumblebee and nucleus of a cell larger than a human. The longest correctly ordered set on the pre-test for non-VR students contained two entities (Panama Canal and Mount Everest). However, the longest correctly ordered set on the post-test contained four entities (Sun, Saturn, Venus, and Mercury). This suggests non-VR students improved their conception of entities at large extremes of scale, which is consistent with an average decrease in variability from 4.28 to 3.79.

Table 4.9 Absolute size ranking for entities prior to and following VR supported instruction.

		VR			
Entities	**Actual Rank**	*Pre-Rank*	*SD*	*Post-Rank*	*SD*
Sun (Diameter)	**1**	2	5.3	1	0.7
Saturn (Diameter)	**2**	3	5.0	5	**5.4**
Venus (Diameter)	**3**	5	3.5	3	**3.8**
Mercury (Diameter)	**4**	4	3.8	2	2.7
Panama Canal (Length)	**5**	8	4.1	8	**6.4**
Mount Everest (Height)	**6**	1	4.6	6	2.7
Golden Gate Bridge (Length)	**7**	7	3.5	4	3.1
Titanic (Length)	**8**	6	3.7	7	2.9
School Bus (Length)	**9**	9	4.0	9	3.7
Human (Height)	**10**	10	3.0	11	**3.3**
Squirrel (Height)	**11**	11	5.0	10	4.4
Bumblebee (Length)	**12**	13	4.7	12	**5.0**
Fruit Fly (Length)	**13**	15	4.4	13	3.4
Human Egg Cell (Diameter)	**14**	20	3.9	15	**4.4**
Skin Cell (Diameter)	**15**	12	4.6	14	**4.7**
Nucleus of a Cell (Diameter)	**16**	18	4.4	16	**5.9**
Influenza Virus (Diameter)	**17**	16	5.1	20	4.3
Ribosome (Length)	**18**	19	5.1	19	5.1
Glucose Molecule (Length)	**19**	17	5.7	17	5.4
Carbon Dioxide Molecule (Length)	**20**	14	6.0	18	4.3

Table 4.10 Absolute size ranking for entities prior to and following non-VR supported instruction.

Entities	Actual Rank	Non-VR *Pre-Rank*	*SD*	*Post-Rank*	*SD*
Sun (Diameter)	**1**	4	6.7	1	1.2
Saturn (Diameter)	**2**	2	4.0	2	2.9
Venus (Diameter)	**3**	5	5.7	3	2.3
Mercury (Diameter)	**4**	1	3.6	4	2.4
Panama Canal (Length)	**5**	7	3.9	8	3.0
Mount Everest (Height)	**6**	8	6.3	6	3.4
Golden Gate Bridge (Length)	**7**	3	3.1	5	**4.5**
Titanic (Length)	**8**	6	2.3	7	**5.3**
School Bus (Length)	**9**	14	3.5	9	3.1
Human (Height)	**10**	10	5.4	12	4.1
Squirrel (Height)	**11**	12	5.1	13	2.1
Bumblebee (Length)	**12**	19	5.1	10	3.8
Fruit Fly (Length)	**13**	20	4.2	15	3.8
Human Egg Cell (Diameter)	**14**	18	5.2	19	3.6
Skin Cell (Diameter)	**15**	15	6.7	14	2.5
Nucleus of a Cell (Diameter)	**16**	13	6.5	11	**7.5**
Influenza Virus (Diameter)	**17**	16	6.1	16	5.7
Ribosome (Length)	**18**	9	5.7	18	**6.3**
Glucose Molecule (Length)	**19**	17	5.3	17	**5.9**
Carbon Dioxide Molecule (Length)	**20**	11	8.7	20	4.7

Relative. Following Delgado's (2013a) procedure, collective landmarks for relative size were calculated prior to and following instruction for both VR (Table 4.11 – 4.12) and non-VR students (Table 4.13 – 4.14). No relative collective landmarks were established among VR students prior to or following instruction. However, following instruction, the Sun and school bus met the criteria for relative collect landmarks among non-VR students.

Prior to instruction and following instruction, 15 entities for VR students and 14 entities for non-VR students met the criteria for accuracy but not precision. This suggests that neither group had strong improvements following instruction. For example, following instruction, five entities among VR students did not meet the criteria for precision or accuracy (i.e., Saturn, Panama Canal, Mount Everest, Human Egg Cell, Carbon Dioxide). Similarly, six entities among non-VR students did not meet the criteria (i.e., Venus, Panama Canal, Mount Everest, Titanic, Influenza Virus, Glucose Molecule). This might be further exacerbated by students' ability to choose a comparison entity (i.e., Earth, Human Hair, or Human) if students' relative conception of the chosen comparison entity was incorrect. For non-VR students, this lack of knowledge is demonstrated in the 8 entities approaching the criteria for an alternative idea (colored gray). However, non-VR students did demonstrate an increase in accuracy among objects in closer proximity to human scale, while VR students did not demonstrate any consistent improvement either near human scale or at the extremes of scale.

Table 4.11 Pretest relative size values for VR participating students from the 25th through 75th percentiles. Correct values for each entity are colored blue with potential alternative ideas colored grey.

Entity	25	30	35	40	45	50	55	60	65	70	75	M
Sun (Diameter)	6	7	7	9	9	9	9	9	9	9	10	8.5
Saturn (Diameter)	2	2	2	3	5	8	8	8	8	10	13	6.5
Venus (Diameter)	3	4	4	4	4	4	4	5	7	8	8	4.7
Mercury (Diameter)	3	4	4	4	4	5	6	6	7	7	7	5.3
Panama Canal (Length)	2	3	3	3	3	3	3	3	3	3	4	3.4
Mount Everest (Height)	2	3	4	6	6	6	6	6	8	9	10	6.0
Golden Gate Bridge (Length)	2	2	2	3	3	4	4	4	5	5	6	4.1
Titanic (Length)	1	1	1	2	2	2	3	4	5	5	5	2.7
School Bus (Length)	-1	0	0	1	1	1	1	1	1	1	1	0.6
Human (Height)	-	-	-	-	-	-	-	-	-	-	-	-
Squirrel (Height)	-3	-3	-3	-2	-2	-2	-2	-2	-1	-1	-1	-1.9
Bumblebee (Length)	-4	-3	-3	-3	-3	-3	-2	-2	-2	-2	-2	-2.3
Fruit Fly (Length)	-5	-5	-5	-5	-5	-5	-4	-4	-3	-3	-3	-4.8
Human Egg Cell (Diameter)	-8	-7	-7	-6	-6	-5	-4	-4	-2	-2	-1	-5.4
Skin Cell (Diameter)	-7	-6	-6	-5	-5	-5	-4	-4	-3	-3	-3	-4.3
Nucleus of a Cell (Diameter)	-9	-8	-7	-7	-7	-6	-5	-5	-4	-3	-2	-5.5
Influenza Virus (Diameter)	-9	-7	-6	-6	-6	-6	-5	-5	-5	-5	-5	-6.0
Ribosome (Length)	-6	-5	-5	-5	-4	-3	-2	-2	-1	-1	-1	-3.5
Glucose Molecule (Length)	-8	-8	-8	-7	-7	-6	-5	-5	-5	-5	-5	-5.1
Carbon Dioxide Molecule (Length)	-8	-7	-7	-6	-6	-5	-5	-5	-4	-4	-3	-5.1

Table 4.12 Posttest relative size values for VR participating students from the 25th through 75th percentiles. Correct values for each entity are colored blue with potential alternative ideas colored grey.

Entity	25	30	35	40	45	50	55	60	65	70	75	M
Sun (Diameter)	8	9	9	9	9	9	9	9	9	9	9	8.1
Saturn (Diameter)	6	6	6	8	9	10	10	10	11	12	13	7.9
Venus (Diameter)	6	7	7	7	7	7	8	8	8	8	8	5.9
Mercury (Diameter)	4	5	5	5	5	5	6	6	6	6	6	6.5
Panama Canal (Length)	0	1	2	2	4	6	6	6	6	6	7	3.7
Mount Everest (Height)	5	5	5	6	6	6	7	7	8	8	8	5.6
Golden Gate Bridge (Length)	2	3	3	3	3	3	3	3	4	4	4	3.2
Titanic (Length)	2	2	2	3	3	3	3	3	4	4	5	3.5
School Bus (Length)	-1	0	0	0	0	0	1	1	1	1	1	0.0
Human (Height)	-	-	-	-	-	-	-	-	-	-	-	-
Squirrel (Height)	-4	-3	-3	-2	-2	-2	-1	-1	-1	-1	-1	-2.4
Bumblebee (Length)	-3	-3	-3	-3	-3	-3	-2	-2	-2	-2	-1	-2.1
Fruit Fly (Length)	-5	-4	-3	-3	-3	-2	-2	-2	-2	-1	0	-1.9
Human Egg Cell (Diameter)	-8	-7	-7	-7	-6	-5	-5	-5	-5	-5	-5	-5.8
Skin Cell (Diameter)	-5	-5	-5	-5	-5	-4	-3	-3	-3	-3	-2	-3.4
Nucleus of a Cell (Diameter)	-8	-8	-8	-7	-7	-7	-6	-6	-5	-4	-3	-5.5
Influenza Virus (Diameter)	-9	-8	-8	-7	-7	-6	-4	-3	-2	-2	-1	-4.0
Ribosome (Length)	-10	-10	-10	-8	-8	-7	-6	-6	-5	-5	-5	-7.3
Glucose Molecule (Length)	-11	-10	-9	-9	-9	-8	-8	-8	-6	-6	-5	-7.9
Carbon Dioxide Molecule (Length)	-9	-8	-8	-8	-8	-7	-6	-5	-5	-5	-5	-6.3

Table 4.13 Pretest relative size values for non-VR participating students from the 25th through 75th percentiles. Correct values for each entity are colored blue with potential alternative ideas colored grey.

Entity	25	30	35	40	45	50	55	60	65	70	75	M
Sun (Diameter)	8	8	8	8	8	8	8	8	9	9	9	8.4
Saturn (Diameter)	3	4	5	6	7	8	8	8	8	8	8	7.8
Venus (Diameter)	2	2	2	3	3	3	3	3	4	4	4	3.6
Mercury (Diameter)	5	6	7	7	8	9	10	11	12	13	14	9.6
Panama Canal (Length)	2	2	2	2	2	2	2	3	3	4	4	1.8
Mount Everest (Height)	1	1	2	2	3	3	3	3	4	4	4	2.8
Golden Gate Bridge (Length)	1	1	1	2	2	2	2	2	2	2	2	4.6
Titanic (Length)	1	1	2	2	3	3	3	3	4	4	4	2.0
School Bus (Length)	0	0	0	0	0	0	0	0	1	1	1	0.4
Human (Height)	-	-	-	-	-	-	-	-	-	-	-	-
Squirrel (Height)	-6	-5	-4	-4	-3	-2	-2	-2	-1	-1	-1	-2.0
Bumblebee (Length)	-3	-3	-3	-3	-3	-3	-3	-3	-2	-2	-2	-3.4
Fruit Fly (Length)	-6	-5	-5	-4	-4	-3	-3	-2	-2	-1	-1	-3.8
Human Egg Cell (Diameter)	-5	-5	-5	-5	-5	-5	-5	-4	-4	-3	-3	-4.2
Skin Cell (Diameter)	-6	-6	-6	-5	-5	-5	-4	-3	-1	-0	1	-3.0
Nucleus of a Cell (Diameter)	-9	-8	-7	-7	-6	-5	-3	-1	0	2	4	-3.0
Influenza Virus (Diameter)	-9	-8	-7	-7	-6	-5	-4	-3	-3	-2	-1	-5.2
Ribosome (Length)	-7	-7	-7	-6	-6	-6	-6	-5	-5	-4	-4	-5.6
Glucose Molecule (Length)	-6	-6	-6	-6	-6	-6	-6	-6	-5	-5	-5	-3.2
Carbon Dioxide Molecule (Length)	-7	-7	-6	-6	-5	-5	-5	-5	-4	-4	-4	-6.0

Table 4.14 Posttest relative size values for non-VR participating students from the 25th through 75th percentiles. Correct values for each entity are colored blue with potential alternative ideas colored grey.

Entity	25	30	35	40	45	50	55	60	65	70	75	M
Sun (Diameter)	9	9	9	9	9	9	9	9	9	9	9	10.2
Saturn (Diameter)	5	5	6	6	7	7	7	7	8	8	8	5.6
Venus (Diameter)	4	4	4	5	5	5	5	5	6	6	6	5.2
Mercury (Diameter)	4	4	4	4	4	4	5	6	6	7	8	6.6
Panama Canal (Length)	3	3	3	3	3	3	4	6	7	9	10	5.8
Mount Everest (Height)	6	6	6	7	7	7	8	8	9	9	10	7.8
Golden Gate Bridge (Length)	2	2	2	3	3	3	4	4	5	5	6	4.0
Titanic (Length)	3	3	3	3	3	3	3	4	4	5	5	6.2
School Bus (Length)	1	1	1	1	1	1	1	1	1	1	1	3.8
Human (Height)	-	-	-	-	-	-	-	-	-	-	-	-
Squirrel (Height)	-2	-2	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1.6
Bumblebee (Length)	-3	-3	-3	-2	-2	-2	-2	-2	-1	-1	-1	-0.4
Fruit Fly (Length)	-4	-4	-4	-3	-3	-3	-3	-3	-2	-2	-2	-3.4
Human Egg Cell (Diameter)	-5	-5	-5	-4	-4	-4	-4	-4	-3	-3	-3	-3.8
Skin Cell (Diameter)	-6	-6	-6	-6	-6	-6	-5	-5	-4	-4	-3	-4.6
Nucleus of a Cell (Diameter)	-7	-7	-6	-6	-5	-5	-5	-5	-5	-5	-5	-5.0
Influenza Virus (Diameter)	-6	-6	-6	-6	-6	-6	-5	-5	-4	-4	-3	-4.2
Ribosome (Length)	-10	-9	-8	-8	-7	-6	-6	-5	-5	-4	-4	-5.0
Glucose Molecule (Length)	-7	-7	-7	-7	-7	-7	-6	-4	-3	-1	0	-4.0
Carbon Dioxide Molecule (Length)	-10	-9	-8	-8	-7	-6	-6	-6	-6	-6	-6	-7.6

Relative size rankings were determined by ranking the mean size values for each entity for VR (Table 4.15) and non-VR (Table 4.16) students prior to and following instruction. Prior to instruction, VR students correctly ranked five entities and non-VR students correctly ranked four entities smaller than the smallest-ranked visible object (fruit fly). After instruction, both VR non-VR students correctly ranked all seven entities smaller than the smallest-ranked visible object (squirrel and fruit fly, respectively). Variability, as measured by the standard deviation, improved on average among both VR (4.28 to 3.79) and non-VR students (4.12 to 3.83).

Table 4.15 Relative size ranking for entities prior to and following VR supported instruction.

		VR			
Entities	**Actual Rank**	*Pre-Rank*	*SD*	*Post-Rank*	*SD*
Sun (Diameter)	**1**	1	4.4	1	2.2
Saturn (Diameter)	**2**	2	7.3	2	7.3
Venus (Diameter)	**3**	5	3.5	4	3.5
Mercury (Diameter)	**4**	4	4.0	3	**4.8**
Panama Canal (Length)	**5**	7	4.6	6	3.8
Mount Everest (Height)	**6**	3	4.2	5	3.5
Golden Gate Bridge (Length)	**7**	6	4.2	8	2.7
Titanic (Length)	**8**	8	3.8	7	1.8
School Bus (Length)	**9**	9	2.7	9	**3.3**
Squirrel (Height)	**10**	10	3.0	12	2.8
Bumblebee (Length)	**11**	11	2.6	11	**3.5**
Fruit Fly (Length)	**12**	14	4.0	10	3.1
Human Egg Cell (Diameter)	**13**	16	5.1	16	3.6
Skin Cell (Diameter)	**14**	13	3.1	13	2.7
Nucleus of a Cell (Diameter)	**15**	17	4.8	15	**5.0**
Influenza Virus (Diameter)	**16**	18	5.9	14	**6.1**
Ribosome (Length)	**17**	12	5.3	18	2.9
Glucose Molecule (Length)	**18**	15	5.1	19	3.8
Carbon Dioxide Molecule (Length)	**19**	15	3.6	17	**5.6**

Table 4.16 Relative size ranking of entities prior to and following non-VR instruction.

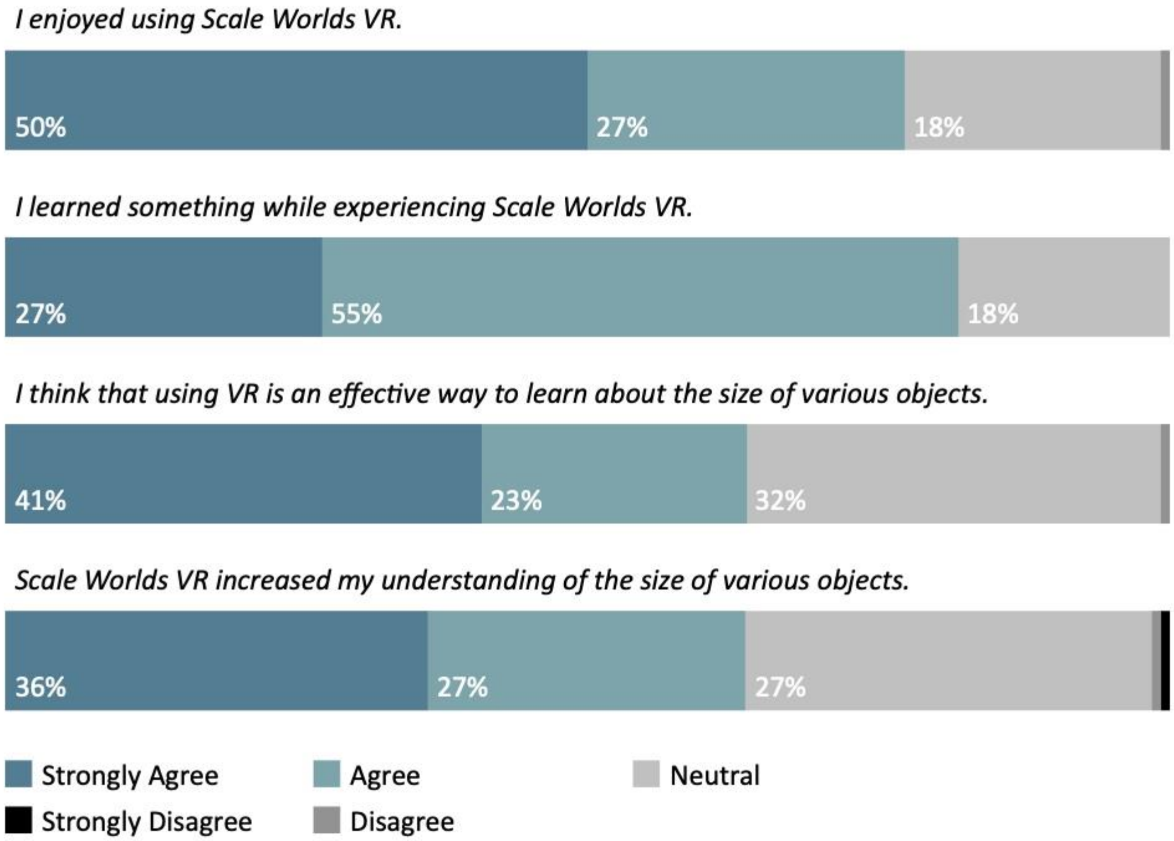
Entities	Actual Rank	Non-VR *Pre-Rank*	*SD*	*Post-Rank*	*SD*
Sun (Diameter)	**1**	2	1.8	1	**3.9**
Saturn (Diameter)	**2**	3	5.7	6	4.6
Venus (Diameter)	**3**	5	2.1	7	1.3
Mercury (Diameter)	**4**	1	5.3	3	5.2
Panama Canal (Length)	**5**	8	3.5	5	**4.3**
Mount Everest (Height)	**6**	6	1.8	2	**2.6**
Golden Gate Bridge (Length)	**7**	4	6.9	8	2.9
Titanic (Length)	**8**	7	3.2	4	**6.7**
School Bus (Length)	**9**	9	3.2	9	**6.3**
Squirrel (Height)	**10**	10	7.1	11	0.9
Bumblebee (Length)	**11**	13	2.7	10	**4.2**
Fruit Fly (Length)	**12**	14	3.1	12	3.0
Human Egg Cell (Diameter)	**13**	15	1.6	13	1.3
Skin Cell (Diameter)	**14**	11	6.0	16	2.5
Nucleus of a Cell (Diameter)	**15**	11	7.1	17	3.1
Influenza Virus (Diameter)	**16**	16	4.8	15	**5.9**
Ribosome (Length)	**17**	17	3.4	17	**6.2**
Glucose Molecule (Length)	**18**	12	6.3	14	4.6
Carbon Dioxide Molecule (Length)	**19**	18	2.5	18	**3.3**

Research Question 1B

What role does Scale Worlds VR play in students' development of size and scale reasoning? (qual)

Feedback Survey. Students' responses indicated that most students enjoyed using SW (77%), learned something while experiencing SW (82%), agree that VR is an effective way to learn about the size of objects (64%), and had an increased understanding of the size of various objects after using SW (63%), see Figure 4.1. When asked to describe something they learned about the size of one object, several (32%) described relative size relationships (e.g., "*how big something can be compared to you*"), one of the five fundamental ways of thinking about size and scale as described by Magaña et al., 2012.

Figure 4.1 Student (n = 22) responses to Likert scale items included on the feedback survey



Statements ranged from a vague reference to size (e.g., "*I learned the different size in between object and how they can compare to us and other objects*") to numerical relationships (e.g., "*that 10 birds sta[c]ked up was equal to a normal human height*"). One student even described the difference between two entities, stating that they "*learned that an ant is 10 times smaller than a bird*." Two references to the serial ordering of objects by size were observed (i.e., "*virus is way smaller than us*", "*I got to know which was bigger or smaller than the other*"). One student described an instance of accommodation, or the Piagetian process of modifying existing mental schemas during new experiences that cannot be assimilated (Inhelder & Piaget, 1958), stating that they "*learned that the size of the mitochondria is a lot smaller than [they] thought it would be*."

Several students (18%) made general references to the use of VR as their favorite thing about SW. Other students described the ability to visualize the size of different entities. For example, one student described that their favorite thing was "*that [they] could see how* *everything related in size and helped [them] understand the difference in size between [them] and things [they] can't see*." The ability to see things beyond the visible describes VR's ability to create the impossible, a previously identified affordance by Steffen et al., 2019. Two students described specific interactive features of the environment as salient (i.e., grabbing, stacking).

While most students enjoyed using SW as previously indicated, the students did identify some barriers or challenges to implementation. The students referenced several general points of tension including inconsistent connectivity, virtual boundaries, and system glitches. For example, one student described their frustration with resetting their height every time they returned to SW. Another student described a difficulty with the virtual boundary, indicating that "*it would not work sometimes when sitting*." However, four students indicated that there was nothing they did not like about SW.

Student Interviews. Interviews from both conditions were analyzed to elicit the role of Scale Worlds VR in students' development of size and scale reasoning using previously identified affordances of virtual reality (Table 3.3) and two emergent codes (Table 4.17). No descriptions regarding virtual reality's ability to diminish negative aspects of the physical world or recreate aspects of the physical world were identified.

Table 4.17 Emergent codes identified through inductive analysis.

Code	**Definition**	**Example(s)**
Interactivity	The ability to act on objects within a virtual environment	*(1) I could pick up things*
Enjoyment	Identified positive affects, or behaviors	*(1) I liked it; (2) it was really fun; (3) I was kind of surprised; (4) more engaged*

Enhance Positive Aspects. Both VR and non-VR participating students indicated that the instructional tool (i.e., Scale Worlds-HMD, Scale Worlds-Web, or Universcale) elevated their learning, suggesting that they provided additional information. For example, a non-VR participating student indicated that "*it helped [them] learn that, like, how many, what numbers next to the 10 is, how many zeros follow it*." Another non-VR student indicated that "*it helped [them] learn the differences between stuff, helped [them] visualize how big something is compared to another thing is*." Similarly, a VR participating student described it as helping "*[them] understand so the numbers…I can memorize it much better.*" Another VR participating student, indicated that the VR helped them learn the powers of 10 "*because [they] saw it face to face*." Similarly, VR participating students repeatedly referenced the ability to "*see it*" or "*visualize*" it. For example, one student indicated that while the teacher is "*talking about what they're trying to teach*" that "*you can actually see it in the VR*." One non-VR participating student, who experienced Scale Worlds following the post-test but prior to the interview, described the necessity of both instructional tools in facilitating meaningful learning:

> *Now, and you're you have a unique experience, because you got to see the VR at the end, and you got to choose this first. Do you feel that you would choose one or the other to help class? Or both? Or neither?*
>
> I'd choose both, because they definitely should see like, like, like the comparison like, like first, like, show them a preview of the comparison, like of how like it would look like. And then, like in the VR, like go to actually it looks like real like you're actually seeing it and you get to you **get a better understanding** of how like it looks like. And I definitely do the VR first because it looked it just looked like it was in real life. Like it just looked better. And like, like it really **helped me understand** like, oh, so this is how big this is. And then when I looked at the blood cells, I was like this is how small it is compared to me and yeah.

A couple of students described VR's ability to enhance other types of content. For example, one student suggested that VR could elevate your reading experience by allowing you "*to see the surrounding of the story and how the characters are*." Another student, described its application in social studies to depict images which you would have "*full control of your hands and head*" to "*look around*" and take "*in a 360 view.*"

Create Aspects that Do Not Exist in the Physical World. Both non-VR and VR participating students described the instructional tool's (i.e., Scale Worlds-HMD, Scale Worlds-Web, or Unvierscale) ability to make entity comparisons (e.g., "*compared to another thing is,* *from all sides to something else*" [non-VR participant], "*I could…measure the bigger thing with them like the bird, I can measure it with the astronaut* [VR participant]"). While one non-VR participating student described the scaling up feature of Universcale, several VR participating students discussed the ability to grow or shrink. For example, one student recommended the use of SW because "*you can like see how big things are and you can go up to their height*." Other students described an egocentric perspective (e.g., "*you can grow like as big or as small you can*", "*we could see how big things were from our point of view*"). Other abilities beyond natural limits the students discussed were the ability to move around in the virtual environment while staying in one place in physical reality, travel back in time, experience different worlds, and look at objects too small to see with the naked eye (e.g., blood cell).

Presence. Not surprisingly, no non-VR students indicated a sense of presence. However, when asked to compare the use of VR as an instructional tool versus a computer, three students described VR's ability to make you feel like you were "*actually in the world*" or "*engulfed*". One student indicated that it elicited a change in perspective making them feel small or big. One student even suggested that if you used only a computer based instructional tool that "*you wouldn't be like experiencing it*."

Interactivity. A few students referenced an interactive element of one of the instructional tools. VR participating students described the ability to "*pick up*" entities. A non-VR participating student described the ability to "*rotate*" 3D images in SW-Web while another referenced the instructional tools in general as "*more interactive*."

Enjoyment. A majority of VR participating students described positive affects towards SW (e.g., "*it was really fun*", "*I liked it*"). Two students described SW's ability to engage students in content (i.e., "*makes certain students more engaged in their learn and not focused on other stuff*") and motivation towards pursing college (e.g., "*it could help students go to college…motivated about school and learning*"). One student described the collaborative process as enjoyable as well stating that when they "*saw it through the iPad*" they "*thought it was pretty cool*" even though they'd already used it before.

Study 2: Exploring Human Impacts on Earth Systems Across Scale Worlds in Immersive Virtual Reality: An Investigation of Students' Integrated Understanding

Research Question 2A

How do middle school students integrate a crosscutting concept and a disciplinary core idea?

For each item of the SPQ and DCI knowledge integration assessment, responses were analyzed using an online plagiarism checker to ensure originality, resulting in a total of 28 (4%) responses overall being omitted (Item 1: six responses; Item 2: two responses; Item 3: no responses; Item 4: four responses; Item 5: three responses; Item 6: two responses; Item 7: three responses: Item 8: two responses; Item 9: three responses; and Item 10: three responses). One student submitted their pre- and post-responses in Spanish, which were omitted as not to misinterpret the student's original meaning from the translated response. Responses were initially coded for the presence of the DCIs for which the items were developed (Table 3.5). However, the open-ended nature of the items led to the identification of additional DCIs (Table 4.18 – detailed codebook in Appendix C). For example, PS1.A was included given references to molecules being indistinguishable (e.g., *its both h20 the same element*) and LS1.A given references to structure or functions within the plant (e.g., *transport nutrients*). Overall, 81% of responses identified at least a single DCI with 7% of responses identifying two (Figure 4.2). Of the 19% not coded, students either did not provide an answer (i.e., *I don't know*), offered overly general responses or were off-task. General responses lacked specificity and could not be attributed to a specific DCI. For instance, one student mentioned that water helps plants grow because *it needs hydrogen*. While this might allude to photosynthesis, *it* appears to refer to water, which is not the reactant requiring hydrogen in photosynthesis. An off-task response when asked about the impact of fertilizer on water quality stated that *it helps the plant grow faster*, neglecting to address its effect on water quality.

Table 4.18 Complete codebook for DCIs including definitions and examples. Additional rules and detailed rationales are included in Appendix C.

Code	Definition	Example(s)
LS1.A Structure and Function	"They have characteristic structures (anatomy and morphology), functions (molecular-scale processes to organism-level physiology), and behaviors (neurobiology and, for some animal species, psychology) (NRC, 2012, p. 143)."	*(1) plants use water* ***to transport nutrients*** *and minerals; (2) water makes the* ***roots like stand up****, hold on to themselves*
LS1.C Organization for Matter and Energy Flow in Organisms	"Plants need water and light to live and grow…Energy needed for life is ultimately derived from the sun through photosynthesis. Plants, algae (including phytoplankton), and other energy fixing microorganisms use sunlight, water, and carbon dioxide to facilitate photosynthesis (NRC, 2012, p. 147)"	*(1) they* ***need water*** *in order to survive; (2) it* ***supplied them with oxygen and hydrogen****, that will then get* ***turning into glucose****; (3) it* ***gives them nutrients***
ESS2.C The Roles of Water in Earth's Surface Processes	"Water is found almost everywhere on Earth, from high in the atmosphere (as water vapor and ice crystals) to low in the atmosphere (precipitation, droplets in clouds) to mountain snowcaps and glaciers (solid) to running liquid water on the land, ocean, and underground. Energy from the sun and the force of gravity drive the continual cycling of water among these reservoirs (NRC, 2012, p. 184)"	*(1) through* ***in the ground*** *because water flows everywhere; (2) crops can* ***get water naturally from rain*** *or get watered by humans*
ESS3.C Humans Impacts on Earth Systems	"Human activities in agriculture, industry, and everyday life have had a major impact on the land, rivers, ocean, and air…and because all of Earth's subsystems are interconnected, changes in one system can produce unforeseen changes in others (NRC, 2012, p. 195)." This includes the "distribution of Earth's water through the modification of streams, lakes, and ground- water (NRC, 2012, p. 194)." "Some negative effects of human activities are reversible with informed and responsible management. For example, communities are doing many things to help protect Earth's resources and environments. They are treating sewage, reducing the amount of materials they use, and reusing and recycling materials (NRS, 2012, p. 195)."	*(1) they impact it by* ***destroying cropland*** *for industrial purposes; (2) Crops can get water naturally from rain or* ***get watered by humans****; (3)* ***By resuing and recycling***

Table 4.18 (continued).

PS1.A	"Pure substances are made from a single type of atom or molecule; each pure substance has characteristic physical and chemical properties (for any bulk quantity under given conditions) that can be used to identify it. (NRC, 2012, p. 108)	*(1) its both h20 the same element*

Figure 4.2 Frequency of DCIs for each assessment item.

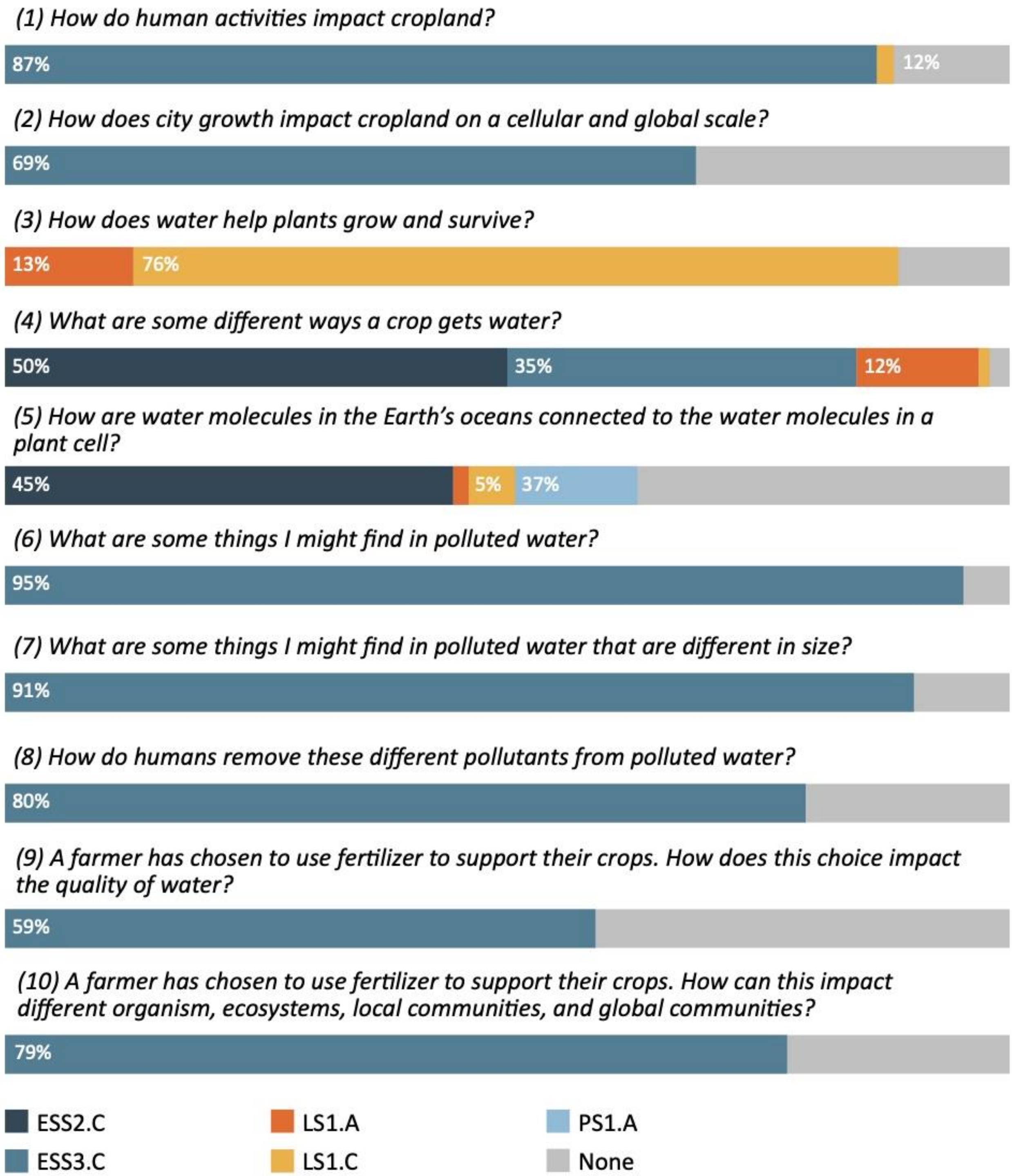


Figure 4.3 Frequency of scale (S); proportion and quantity (PQ); and both (S+PQ) for each assessment item.

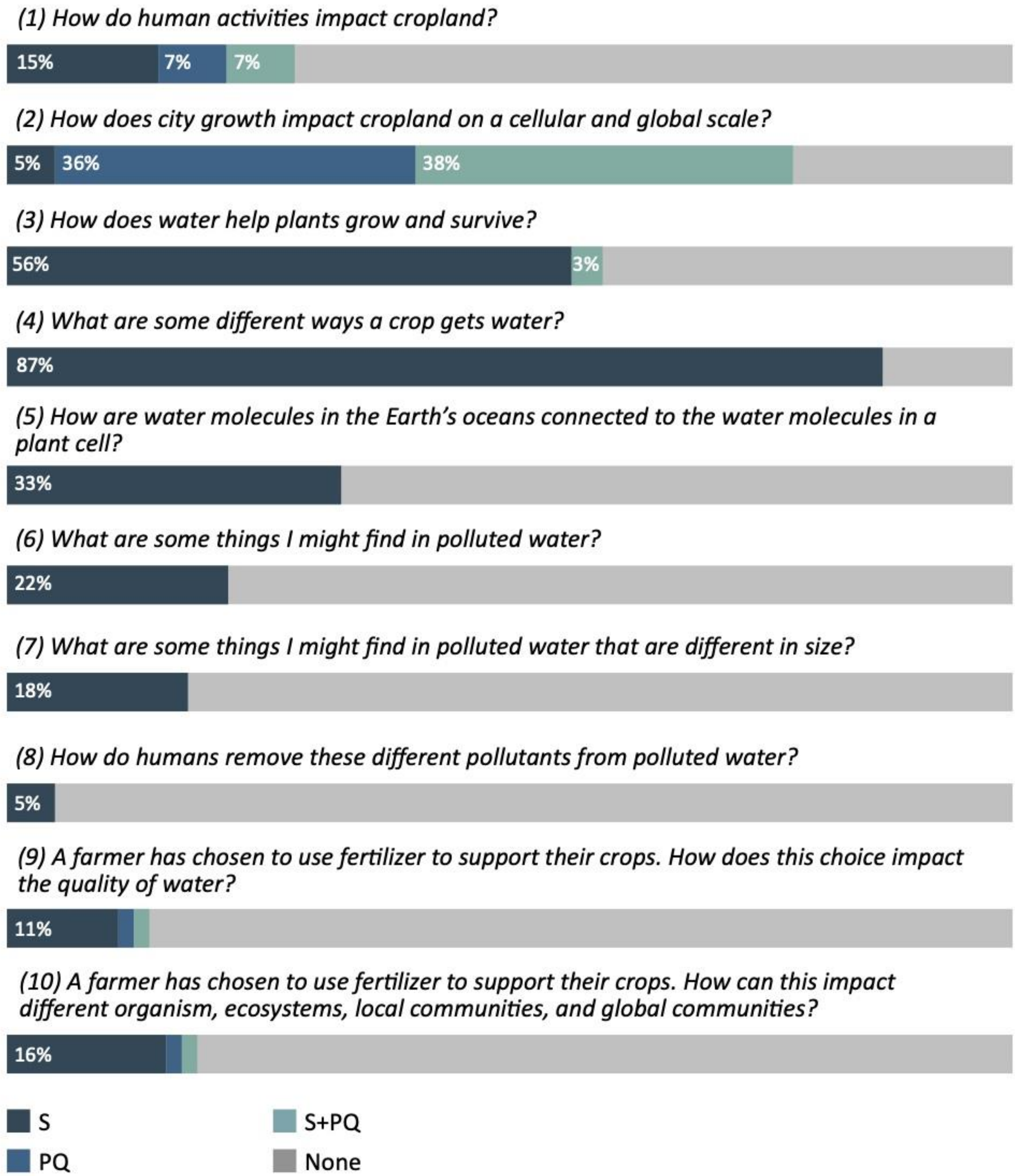
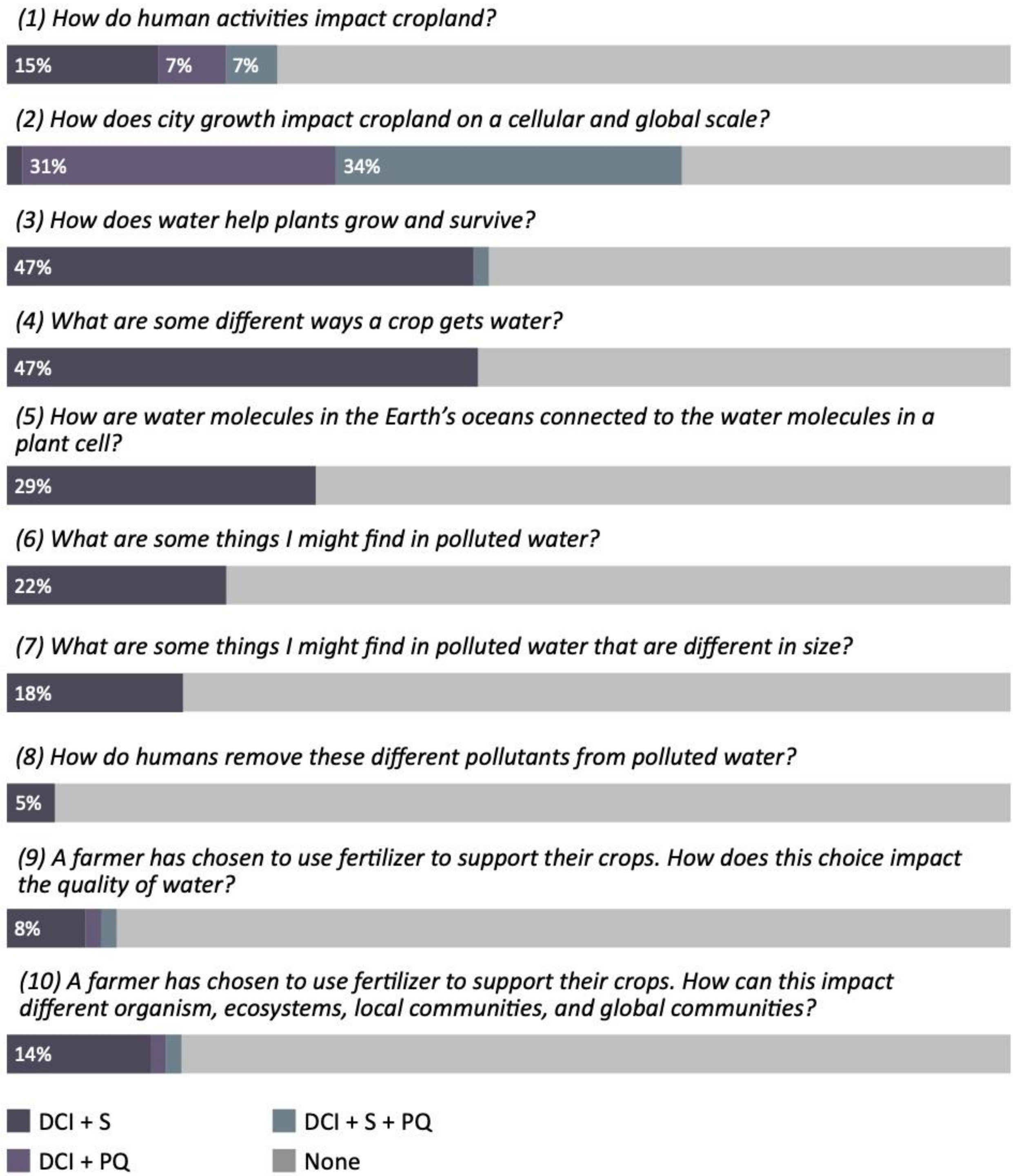


Figure 4.4 Frequency of integrated responses for each assessment item.



Next, each response was coded for the presence of SPQ as discussed previously and detailed in Appendix D. Overall, 36% responses (Figure 4.3) were coded as having the presence of either scale (26%), proportion and quantity (5%), or both (5%). However, most students did not indicate SPQ, either focusing on a single scale (e.g., *water helps the plant keep hydrated and stay alive. They need water in order to survive* – only discussing plant or human scale), not describing quantitative variables (e.g., *humans improve cropland* – students did not report a quantifiable variable like increased access to quality water or higher crop yields), or providing descriptions that lacked a sense of scale (e.g., *they have to save water*). Responses coded for a DCI and SPQ (Figure 4.4) were then analyzed to identify themes within and across items.

Integrating DCIs and SPQ within items. For item 1, a few students applied scale to express the magnitude of an impact. For example, one student described the role of farmers in providing "*food for the people on earth.*" Another student described the extent of catching crops on fire by referencing the recent "*wildfires in Hawaii*." The actions identified by three students indicated scale (e.g., *genetically modifying plants* or *urbanizing areas*), while other students only made references to other scales (e.g., *sunlight*, *air*). However, four students did describe the response of one variable (i.e., amount of cropland) to an increase in urbanization (e.g., building houses). Three students integrated SPQ by describing the response of one variable to changes in another across scale contexts. For example, a student described how the growth of the human population (big scale) results in an increase in the number of crops we must plant (human scale).

Item 2, which asked a similar question but sought to elicit more instances of integration, led to a 41% increase in integrated responses. The item asked students to describe a response or effect related to a quantitative change in a variable (i.e., city growth). Therefore, unsurprisingly, most students described a quantitative response (e.g., *more people…more cropland; more pollution…less food*) with few discussing variable responses across scales (e.g., *bigger population…more food to plant*). Students tended to take a unidirectional approach, only discussing a decrease in cropland as a response to growing cities. However, one student did express the need for *some sacrifices…like cutting off some of the land needed for food and* *make it for houses*. Only one student discussed impacts at the microscale, indicating that the pollutants might impact the soil causing *some microorganisms to stop appearing*.

For item 3, most students only referenced objects (e.g., *nutrients*) or processes (e.g., *photosynthesis*) that occur at scales beyond human. For example, there were limited responses that indicated that students were thinking about photosynthesis as a molecular process (e.g., *it supplies them with oxygen and hydrogen, that will then get turned into glucose*) with most just citing the process itself (e.g., *photosynthesis helps it grow faster and survive more longer*). While photosynthesis requires input energy from the Sun, a reference to scales larger than human, only three students mentioned it. Most students discussed water's role in providing nutrients with some expanding on this by describing its role in "*carry[ing] nutrients throughout the plant*."

Students were also asked to indicate the ways in which plants acquire water (item 4) to elicit integration among the various scales at which water interacts, moves, or is stored. However, most only indicated rain, or rainfall, with some considering absorption through the roots. This indicates that students are considering scales beyond human scale. Yet, descriptions of systemic interactions across scales were not observed (e.g., *rain either on the plant or transported via rivers or through water moving underground*). When asked specifically about the connections (item 5) among water sources at different scales (i.e., water molecules in a plant cell and the ocean) no increase in integration was observed. However, some students did describe connections across scale (e.g., *water from the ocean evaporates it goes through the water cycle causing the rain to give the plants water*) but descriptions were underdeveloped (e.g., *rain because the water cycle; plants use the earth's water*). One student applied the properties of matter to make connections across scales, indicating that *water is made of H2O in the Earth's oceans and so is the water we use to water the plants.* Similarly, one student described it as *water is water whether it's from the earth or our sink it still gives plant nutrients.*

Item 6 asked students to identify water pollutants. Students typically only listed pollutants at a single scale, either human scale (e.g., *trash*) or the microscopic scale (e.g., *germs*). Only 14 students identified pollutants across scales (e.g., *trash, dirt, bacteria, food*). No student identified or described any pollutants larger than human scale. However, when asked to list pollutants of different size (item 7), only 11 students indicated pollutants across scales. Most students listed additional objects at human scale that varied only by a order-of-magnitude or less in size. However, two students did indicate items larger than human scale (i.e., oil spills). Although, no student indicated pollutants across human, microscopic, and big scale.

For item 9, students were asked to describe the impact of fertilizer on water quality. However, few students considered the impact beyond the plant itself (i.e., *helps the water be better for the plant*) with only three students considering impacts across scales like contaminating the water supply or sea. Similarly, for item 10, in which explicit impacts across scale were sought, only four students identified impacts such as contaminating the water supply or limiting communities access to *healthy* water. One student did consider the economic impact by describing how by growing plants faster people will be able to make more money.

Integrating DCIs and SPQ across items. Across items, 32% of total responses were identified as integrated, or coded for both a DCI and S, PQ, or both (Figure 4.4). Analysis of integrated responses revealed that explicit prompts were necessary for SPQ integration. For instance, students more readily integrated SPQ when explicitly prompted, as seen in item 2 of the integration assessment, which asked about the impact of city growth on cropland. Responses tended to be unidirectional, focusing on immediate effects rather than dynamic discussions of reciprocal impacts but demonstrated an awareness of long-term and large-scale impacts of local processes, aligning with the Framework for K-12 Science Education (NRC, 2012). For other implicitly and explicitly paired items, only subtle changes in the frequency of integrated responses were observed. For example, item 5 sought to elicit an increase in integrated responses that might not have been achieved with item 4 but resulted in a 16% decrease in integrated responses. However, some richer descriptions of integration were elicited. For example, integration within item 4 was limited to references to multiple scales (e.g.*, basins, humans can give water, rain*) while item 5 elicited descriptions of systemic interactions across scale (e.g., *water that is in the ocean is used for rain which is then drop on soil where plants are allowing them to be absorbed into the plant*).

Across all items students tended to focus on objects or processes at a single scale. For instance, students only identified pollutants at human scale, not considering microscopic scale even when prompted to identify pollutants of difference sizes. In addition, references to scale primarily consisted only of references to other scales with underdeveloped descriptions such as only referencing photosynthesis, rain, or sunlight without discussing interactions at various scales (e.g., *water molecules are used to create glucose molecules which the plant breakdown to produces energy and grow*).

Research Question 2B

How does immersive virtual reality-supported science instruction influence middle school students' integrated understanding of a crosscutting concept and a disciplinary core idea?

SPQ and DCI Knowledge Integration Assessment. A comparison between VR and non-VR integrated responses following instruction did not reveal any consistent patterns differentiating the two groups. A chi-square test of independence showed that there was no significant association between the use of VR and the frequency of integrated responses overall, χ^2 (1, N = 330) = 0.3, p = .58, integrated responses per type, χ^2 (2, N = 119) = 1.6, p = .45, or across items, χ^2 (9, N = 119) = 4.7, p = .86. Therefore, this suggests that the VR did not increase students' integration of DCIs and SPQ as originally hypothesized.

Student Interviews. Interview transcripts were reviewed to explore the activities that students identified as supporting their integrated understanding. Of the 11 participants, all but one student was asked to identify activities following a discussion around the integration of a CCC and DCI. Most students (5) did not initially identify any activity. However, aerial photography, SW-HMD, and the poster (i.e., the physical scale students developed on poster paper) were each identified twice. SW-Web and videos were each identified once. When probed again regarding what stood out to the participants, one student identified the poster, and one student identified content (e.g., *how we should like not waste food or water*). One student, when asked what they picture when a plant is growing, described timelapse videos. Another student indicated the *changing my size in* SW-HMD was one thing they would remember. One student indicated identified a video in class that discussed the challenges faced by farmers from underrepresented and underserved populations as improving their understanding. No other activities were highlighted by the students during the follow-up interview. While students did identify SW-HMD as a salient activity, no students explained its role in fostering integrated knowledge, suggesting the need for further data regarding the role of VR in establishing integrated knowledge.

CHAPTER 5: Discussion

Study 1: Investigating Middle School Students' Scale Reasoning Through Immersive Virtual Reality-Supported Science Instruction

Research Question 1A

How do middle school students' size and scale reasoning change after VR-supported science instruction? (QUAN)

The co-developed nine-day instructional unit integrating human impacts with the crosscutting concept of "scale, proportion, and quantity" (NGSS Lead States, 2013) led to statistically significant learning gains among both VR and non-VR supported instruction as demonstrated by an increase in students composite score on the Assessment of Size and Scale Cognition (ASSC). This demonstrates the ability of instruction with minimal additional resources to support students' size and scale cognition. Specifically, VR-supported instruction significantly improved students' performance on ordering, relative, and absolute task items. Previous interventions were limited in scope (e.g., Chesnutt et al., 2018; Jones et al., 2007) or were unable to achieve significant learning gains across these three fundamental ways of thinking of size and scale (e.g., Magaña, 2014; Delgado et al., 2015). Non-VR students descriptively improved on ordering, relative, and absolute task items with medium to large effect sizes (Table 4.2). When comparing across groups, VR students' composite scores were descriptively higher (+6.2 pts) with a medium effect size, but results were not statistically significant. This might be a result of low statistical power given the small sample of paired responses among students who received non-VR instruction. Out of 18 students who originally assented, I was only able to obtain eight matched pairs which was reduced to five when removing responses for lack of quality as measured by item duration. While the results might suggest an effect it does not provide a clear understanding of learning gains achieved by both groups. Therefore, I discuss analyses and results for each section separately while integrating quantitative and qualitative findings.

Ordering. It was hypothesized that the direct embodied experiences with entities at the extremes would support students' conceptions of size and scale (Wilson, 2002). However, average ranking revealed that the small number of entities correctly ordered were within 10^3 m to 10^{-3} m, which I refer to as human scale based on previously identified boundaries (Tretter et al., 2006a; Trate et al., 2019). This is consistent with prior findings, which indicate that oneself is a major reference point (Tretter et al., 2006a). Even though ordering accuracy only slightly improved among both groups, students demonstrated increased awareness of microscopic entities. Younger groups of students have been known to identify a visible object as the smallest they can think of (Tretter et al., 2006a). For example, Waldron et al. (2006) found that middle schools often identify a macroscopic entity (e.g., grain of salt) when asked to identify the smallest thing they can think of. Results indicate instruction supported students' in increasing their awareness of things too small to see. Students participating in the VR supported instruction correctly ranked all seven microscopic entities smaller than the smallest-ranked visible object with non-VR students correctly ranking four. However, several students across groups identified the physical scale they developed in class during the follow-up interviews as contributing to their conception of scale. Therefore, it is unclear which instructional aid was supporting students' development of a mental model of scale.

Delgado et al. (2015) demonstrated that both public and private middle school students on average ranked the cell larger than an atom, but molecule, mitochondrion, or virus smaller than atom. However, both VR and non-VR students did not demonstrate such performance with VR students on average ranking a molecule larger than a cell, nucleus, and virus, and non-VR students on average interspersing macroscopic entities among submacroscopic entities. This suggests that the integrated instructional unit was unable to achieve similar performance metrics as reported by Delgado et al. (2015). This might be attributed to the transfer entities included within the ASSC (Table 2.3), which were similar as to establish a mental link to those included in SW and the unit, but different as to not assess students recall of entities. Therefore, students might have been presented with a challenge when attempting to order the objects based on size that was not present in previous studies (Delgado et al., 2015).

Absolute. Delgado (2013a) defined a collective landmark as an anchor common among a group of learners that instructors can use to assist students in establishing new landmarks. Collective landmarks were developed for events to assist students in navigating deep time, which is always relative to the present time. However, the size of an entity can be absolute or described relative to a reference entity. Therefore, collective landmarks were assessed for both absolute and relative size values. However, the only absolute size collective landmark identified was the Sun among students who participated in VR-supported instruction. This suggests that the absolute size of the Sun (10^9 m) might be an anchor upon which the instructor could build. For example, supporting the student in establishing that molecules are roughly 10^{-9} m. This could be further supported if an absolute collective landmark could be established at the human scale (10^{-3} m to 10^3 m), fostering a link between everyday experience and entities at the extremes of scale. It is not surprising that only limited absolute collective landmarks could be established as research suggest that younger children may clump entities as smaller or larger relative to other things, developing absolute size of things much later (Jones et al., 2007; NRC, 2012).

While no alternative ideas, when the accepted value falls within the 25th and 75th percentile but the 25th and 75th percentile estimates are not within an order-of-magnitude, were detected, it should be noted that among both groups, greater than 30% of students imagined the fruit fly an order-of-magnitude larger. Interviews with students indicated that this might be due to lack of familiarity, with two students indicating that they had never seen them. This might have resulted in students reporting the size of a common housefly or horsefly, both of which are one order-of-magnitude larger. Similarly, the Panama Canal did not meet criteria for either precision or accuracy, again indicating a lack of familiarity. One student elaborated on this during their interview describing the canal as *something inside the body,* like your colon. Therefore, this demonstrates that Delgado's (2013a) method for establishing collective landmarks might not only support instructors in identifying common anchors for instruction but also highlight a lack of familiarity with others. This is particularly salient given that conceptions of scale are linked to direct experiences, as demonstrated by Tretter et al. (2006a) and Chesnutt et al. (2018), which might be unknown to the instructor.

A decline in accuracy among students participating in VR-supported instruction was observed for several entities (e.g., bumblebee, skin cell, virus). This could be attributed to the dynamic nature of the ASSC. Absolute task items are presented either with the size given to identify the entity or with the entity given to identify the size. The method of assessing an entity depends on the students' order from the previous section. Therefore, it is possible that entities were assessed differently on the pre-test and post-test, potentially resulting in varied accuracy. However, further research is needed due to the novelty of the ASSC. Conversely, only a few entities among non-VR students were observed to decrease in accuracy (Table 4.10), which might suggest that something within the virtual environment is introducing misconceptions, although none were identified during the follow-up interviews.

Relative. Prior research has demonstrated that estimating relative size is difficult and resistant to change (Delgado et al., 2015). However, the results demonstrate that both groups showed learning gains regarding relative size. While a significant improvement in relative size task items was observed among students participating in VR-supported instruction, only the non-VR students successfully demonstrated the establishment of relative collective landmarks (i.e., the Sun and school bus). However, students participating in VR-supported instruction came close to achieving a relative collective landmark for the Sun. It might be suggested that since the Sun is at the extreme, determining its relative and absolute size is easier. However, during the development of the ASSC (Harper-Gampp et al., 2024a, 2024b, 2024d) this was taken into consideration by ensuring the range of options extended at least two orders-of-magnitude beyond the correct value.

Based on the progression with which learners acquire a concept of scale (NRC, 2012), it was hypothesized that students would more readily establish relative collective landmarks. However, no significant improvement was noted following instruction, corroborating previous findings regarding its resistance to change (Delgado et al., 2015). For instance, several entities among students participating in non-VR supported instruction were approaching the criteria for an alternative idea. This suggest that the students were unable to establish clear distinctions, possibly due to a lack of understanding regarding multiplicative relationships (i.e., three times more, three times less; Vergnaud, 1988). Nonetheless, students from both groups collectively ranked all seven microscopic entities smaller than the smallest-ranked visible object, demonstrating an increase in awareness of microscopic objects. In addition, the salience of relative size was echoed throughout the feedback survey with a majority of students (32%) describing a relative size relationship when asked to describe something they learned (e.g., *an ant is 10 times smaller than a bird* or *how they compare to us*). No student indicated a reference to absolute size as something they learned. This corroborates previous findings regarding the lack of precision or need for specific spatial metric information among everyday language (Tretter et al., 2006a).

Tretter et al. (2006a) calculated the average absolute ranking among elementary, middle school, high school, gifted seniors, and experts. The standard deviation was defined as the variability in students' responses, which they observed declined from younger to more experienced learners. While the specific entities assessed are different, I sought to situate my results among theirs by determining the average variability across groups. Both groups variability decreased following instruction, suggesting an improvement in precision. In addition, the average variability among both groups prior to instruction fell between the average variability identified by Tretter et al. (2006a) for elementary students and middle school students. However, following instruction both groups were approaching the variability identified for high school students. This might suggest that the instructional intervention not only improved students' conception of absolute size but assisted in closing the gap among them and other middle school students from the central region of North Carolina.

Research Question 1B

What role does Scale Worlds VR play in students' development of scale reasoning? (qual)

Students engaged in the VR-supported instruction not only achieved statistically significant learning gains but also perceived increased learning gains after completing the co-developed nine-day instructional unit. The majority (82%) reported learning while using Scale Worlds (SW), with 63% indicating that SW enhanced their understanding of entity sizes. In follow-up interviews, non-VR students expressed similar sentiments about web-based tools, stating that either SW-Web or Universcale "*helped them learn*" or "*helped [them] visualize how big something is compared to another*." Both groups showed interest in entities too small to see, aligning with observations by Jones et al. (2007). In addition, students participating in VR-supported instruction consistently described the embodied nature of SW. For instance, many described the ability to alter their perspective by shrinking or growing (e.g., *you can grow* or see entities *from our point of view*). This highlights SW's capacity to facilitate embodied interactions across a wide range of scale, inaccessible for direct experience due to either being too small to see or too large. This supports earlier assertions (Tretter et al., 2006a; Jones et al., 2008; Magaña, 2014) about VR's potential for facilitating interactions at extremes of scale. While web-based tools allowed scaling and rotation, this aspect was not consistently emphasized by participants. Conversely, students described embodied interaction with SW, such as picking up objects and measuring size comparisons (e.g., *pick up things and measure the bigger thing with them like the bird…I can measure the astronaut*). This underscores VR's ability to create novel experiences impossible in reality, an affordance identified by Steffen et al., 2019. Several students highlighted how the instructional tools aided their learning, particularly in visualizing content like scientific notation. Many students mentioned the capability of comparing the sizes of different entities, a skill that Magaña (2014) proposes could enhance students' proficiency in organizing and categorizing objects. While both groups referenced comparing entity sizes, VR students specifically noted the immersive visual experience, suggesting VR's advantage. However, it is evident that both tools support scale reasoning. One student, who experienced SW after the post-test but before the interview, reinforced this by advocating for both tools in effective instruction. They viewed SW as a scaffold for connecting the web-based tools and making relative comparisons (e.g., *red blood cells…how small it is compared to me*).

As expected, given the novelty of integrating VR into traditional classroom settings (Matovu et al., 2023; Luo et al., 2021), several barriers to implementation were identified, particularly considering the scale of deployment. One significant challenge highlighted in feedback was virtual boundaries, which I personally observed during implementation due to space limitations and the need to rotate devices among students. Additionally, since VR devices are typically designed for individual use, accommodating multiple users posed difficulties, as evidenced by the limited number of accounts permitted on each device. This aspect also causes issues in SW, requiring students to recalibrate their height frequently to ensure accurate representations of entities and to reenable casting after each rotation. Therefore, findings suggest that future implementations not only address general technological barriers but also consider the specific deployment context, traditional classroom setting, compared to the device's intended us. However, a potential solution to many of these challenges could involve incorporating collaborative casting, which I found helped address student concerns and has been previously recognized as a method to enhance equity in the VR experience (Harper-Gampp et al., 2023c).

Study 2: Exploring Human Impacts on Earth Systems Across Scale Worlds in Immersive Virtual Reality: An Investigation of Students' Integrated Understanding

Research Question 2A

How do middle school students integrate a crosscutting concept and a disciplinary core idea?

The integration of the crosscutting concept (CCC) of "scale, proportion, and quanity" (SPQ) with various disciplinary core ideas (DCIs) was observed among 35% of total responses, with scale being the predominant focus. This prevalence of scale instead of proportion and quantity might be a result of Scale World's (SW) strong alignment with the FS2C (Magaña et al., 2012a) and its goal of establishing embodied interactions with entities across 20 orders-of-magnitude. Students utilized scale to demonstrate the magnitude of human impacts, an application for which the Framework for K-12 Science Education explicitly identifies, indicating the importance of recognizing how local processes can have large-scale impacts (NRC, 2012). In addition, when explicitly prompted, students applied proportionality to explain how the growth of cities impacted the amount of cropland and food. This demonstrates the ability of SPQ in supporting students in explaining and predicting the impact of phenomena aligned with a DCI, as identified by Nordine and Lee (2021) regarding the role of CCCs. However, many students only referenced other scales (e.g., *rain or by hand*) without establishing clear comparisons or connections across scales. This might result from the implicit nature of CCCs within science instruction, leading to students being unable to engage meaningfully in this type of science learning (Olson and Loucks-Horsley, 2000). For instance, this study took place in the state of North Carolina, which has yet to adopt the NGSS or incorporate CCCs within state science standards (North Carolina Department of Public Instruction, 2024).

In addition, many responses did not explicitly incorporate scale, often focusing solely on a single scale (e.g., human scale), particularly the scale referenced within the question stem. For example, when asked about the impact of human activities on cropland, responses predominantly centered on the human scale (10^{-3} m to 10^{3} m), with only a small fraction referencing larger (>10^{3} m) or microscopic scales (10^{-3} m). However, previous studies (e.g., Tretter et al., 2006a; Trate et al., 2019), have highlighted our tendency to use our bodies as reference points from early developmental stages through our professional careers (Jones & Taylor, 2009). This tendency was further evidenced in students' identification of pollutants. When asked to identify pollutants of different sizes, students predominantly listed macroscopic entities that varied by no more than an order-of-magnitude. This underscores the significance of self-reference in determining scale, a concept identified by Tretter et al. (2006a) as a key conceptual boundary from elementary students to experts. However, this reliance on self-reference might have been constrained by the imprecision inherent in everyday language. Tversky (1993) suggests that students construct spatial mental models categorically, lacking metric information. Everyday language, characterized by imprecise phrases like "next to" or "between," further complicated the interpretation of student responses. For instance, several students identified chemicals as pollutants, but they were not categorized as microscopic since students may have been thinking of chemicals ranging from macroscopic (e.g., a barrel of oil) to microscopic (e.g., molecules of PFAs).

While results suggest an increase in students' awareness of microscopic entities, there were very few descriptions of the microscopic scale across all items, such as in discussions of genetically modified plants. The scarcity of microscopic scale descriptions may indicate that students lacked sufficient experience necessary, as suggested by Paivio (1990), for processing by the cognitive system. Exploring the microscopic world requires access to tools that students may not have, especially outside of school (Jones et al., 2021), which might have resulted in an overall limited number of experiences with such worlds. Therefore, further experience is needed before microscopic scales can be effectively integrated into disciplinary content.

A key role of CCC is providing multiple lenses for exploring DCI-related phenomena (Nordine & Lee, 2021). However, this might have been hindered by students' underdeveloped conceptions of key concepts such as photosynthesis, the water cycle, and plant function and development. For example, most students only identified photosynthesis as important in making food without describing the microscopic or chemical nature of the process. However, an understanding of these *hidden dimensions*, those too large or too small to experience naturally, is critical in developing a systems-level thinking of processes like photosynthesis or the water cycle (Lee et al., 2019). Yet, extant literature supports students' difficulty in establishing these connections. For example, Hmelo et al. (2000) reported 6th graders' similar difficulties differentiating microscopic and macroscopic processes regarding the human respiratory system.

Research Question 2B

How does immersive virtual reality influence middle *school students' integrated understanding of a crosscutting concept and a disciplinary core idea?*

When comparing across groups, no distinct patterns were identified. Students demonstrated similar linguistic ambiguity and underdeveloped conceptions of key concepts regardless of condition. A difference in the frequency of integrated responses was also not detected. However, this might be a result of instructional tools focusing solely on size and scale reasoning. The virtual environment itself does not situate size and scale reasoning within disciplinary content, which might better support students integrated understanding. This instructional unit relied on curriculum to depict scale stories, establishing connections between content and the instructional tool. However, future implementations might embed scale stories within the virtual environment. For example, a teacher who previously developed curriculum with SW described a storyline following the life cycle of a plant across scales (Harper-Gampp et al., 2024c). They suggested experiencing a large field and then scaling down to watch a pollinator interact with the reproductive parts of the plant. This was also made evident in the activities identified, in which students either made no identification or only made general references to activities, not specific elements that supported there understanding. For example, two students identified the aerial photography activity as supporting their learning, but it is unclear what aspect of the activity itself supported the students' integrated understanding. Therefore, this indicates a need for further research regarding the role of VR in supporting students' integration of DCIs and SPQ.

Limitations and Delimitations

The novelty of implementing VR within authentic K-12 learning environments presented challenges that might have increased the extraneous load experienced by students' resulting in deviations from content. Future research should consider having students complete an unrelated activity in VR prior to increase comfortability and decrease the novelty effect. Without incentives for students, it was difficult to maintain consistent attendance which resulted in less than expected paired responses, limiting the power of analyses. Investigation was isolated to eighth grade students' integration of a subset of DCIs and SPQ. Therefore, results might not be generalizable to other CCCs, DCIs or other grade levels. Given the linguistic ambiguity of students' responses it might be valuable to explore other methods for capturing students' integration.

Conclusion

The enactment of the co-developed nine-day instructional unit, integrating human impacts with the crosscutting concept of "scale, proportion, and quantity" (NGSS Lead States, 2013), resulted in statistically significant learning gains among both VR and non-VR students. VR-supported instruction led to statistically significant improvements in students' performance on ordering, relative, and absolute task items, showcasing the ability of instruction with minimal additional resources to enhance students' conception regarding these aspects, which previous interventions failed to achieve (e.g., Magaña, 2014; Delgado et al., 2015). When comparing across groups, VR students performed descriptively better with a medium effect size, but the results were not statistically significant.

However, a few nuanced differences emerged between the two instructional approaches. Results indicated that both instructional techniques notably improved students' awareness of the microscopic world, with students collectively ranking more microscopic entities smaller than the smallest-ranked visible object following instruction. However, only the Sun and school bus were identified as collective landmarks for relative size among non-VR students, with VR students not achieving any collective landmarks. Collective landmarks, which serve as instructional anchors, might support instructors in establishing new landmarks regarding scale. Nonetheless, further research is required to explore this aspect. Additionally, the identification of possible alternative ideas (e.g., fruit fly, Panama Canal) was corroborated with interview data, suggesting the promise of this technique for highlighting a lack of familiarity, particularly given that concepts of scale are linked to direct experiences, which might be unknown to the instructor.

Students integrated the CCC of SPQ with various DCIs to demonstrate the magnitude of human impacts on the environment, aligning with educational frameworks (NRC, 2012). However, assessment items aimed at eliciting integration did not consistently result in increased integration. Instead, students tended to focus on a single scale or only reference other scales, without describing connections across scales. Underdeveloped conceptions of key concepts might hinder their ability to assess concepts through multiple lenses (Nordine & Lee, 2021), a critical skill in developing systems-level thinking of processes like the water cycle (Assaraf & Orion, 2005).

While a statistically significant difference was not obtained, findings suggest that SW established a positive learning experience and facilitated embodied interactions with entities at the extremes of scale, supporting students in visualizing abstract concepts. However, barriers to implementation, such as technical limitations and logistical challenges, underscore the importance of careful planning and consideration when integrating VR into classroom settings. This dissertation serves as a framework to support future instructional interventions in addressing these challenges while capitalizing on the affordances of VR technology to provide immersive and engaging learning experiences. Collaboration with casting and other techniques to promote equitable access and meaningful interaction with VR environments may help to overcome implementation barriers and maximize the benefits of this innovative instructional approach.

Overall, this study contributes valuable insights into the potential of VR-supported instruction in fostering students' size and scale reasoning and underscores the importance of ongoing research and development in educational technology integration. When examining the influence of immersive VR on integrated understanding, no distinct patterns were identified across groups. However, results contribute to ongoing research seeking to understand the role of CCCs in understanding scientific phenomena. Further research is needed to explore instructional strategies that effectively support the integration of SPQ with DCIs, fostering a deeper understanding of complex scientific concepts among students. Nevertheless, the findings highlight the potential value of VR-supported instruction in enabling embodied interactions with entities across a wide range of scales.

Implications

While the unit did require a considerable amount of instructional time given the integration of SPQ within the context of multiple DCIs, this demonstrates the possibility of achieving learning gains across three of the fundamental ways of thinking of size and scale with limited resources. This dissertation serves as a guide for future investigations into the application of VR within traditional classroom settings, highlighting the potential of VR to establish embodied interactions. However, results suggest that the way VR supports instruction is nuanced, not clearly better or as effective as using other instructional technology. Similarly, findings demonstrated the challenges in identifying students' integration of DCIs and CCCs given the lack of precision among everyday language. This dissertation contributes to the lack of research on when and how to use VR to support middle school science instruction (Matovu et al., 2023) and contributes to the field's research on interactions between CCCs and DCIs.

REFERENCES

Aldas, K., & Yapici, R. (2014). Investigation of effects of scale and surface roughness on efficiency of water jet pumps using CFD. *Engineering Applications of Computational Fluid Mechanics*, *8*(1), 14–25. https://doi.org/10.1080/19942060.2014.11015494

Altin, M., Schwartz, Z., & Uysal, M. (2017). "Where you do it" matters: The impact of hotels' revenue-management implementation strategies on performance. *International Journal of Hospitality Management*, *67*, 46–52. https://doi.org/10.1016/j.ijhm.2017.08.001

American Association for the Advancement of Science (AAAS). (1989). *Science for all Americans*. New York: Oxford University Press.

American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Ashfaq, A., Khursheed, N., Fatima, S., Anjum, Z., & Younis, K. (2022). Application of nanotechnology in food packaging: Pros and cons. *Journal of Agriculture and Food Research*, *7*, 100270. https://doi.org/10.1016/j.jafr.2022.100270

Assaraf, O. B.-Z., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching*, *42*(5), 518–560. https://doi.org/10.1002/tea.20061

Attoh-Mensah, E., Loggia, G., Schumann-Bard, P., Morello, R., Descatoire, P., Marcelli, C., & Chavoix, C. (2020). Adverse effects of anticholinergic drugs on cognition and mobility: Cutoff for impairment in a cross-sectional study in young-old and old-old adults. *Drugs & Aging*, *37*(4), 301–310. https://doi.org/10.1007/s40266-019-00743-z

Bibic, L., Druskis, J., Walpole, S., Angulo, J., & Stokes, L. (2019). Bug off pain: An educational virtual reality game on spider venoms and chronic pain for public engagement. *Journal of Chemical Education*, *96*(7), 1486–1490. https://doi.org/10.1021/acs.jchemed.8b00905

Bourdieu, P. (1977). *Outline of a theory of practice* (Vol. 16). Cambridge University Press.

Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste*. Harvard University Press.

Brook, A., Briggs, H., & Driver, R. (1984). *Aspects of secondary students 'understanding of the particulate nature of matter*. Children's Learning in Science Project, Centre for Studies in Science and Mathematics Education, the University.

Brown, N. R., & Siegler, R. S. (2001). Seeds aren't anchors. *Memory & Cognition*, *29*(3), 405–412. https://doi.org/10.3758/BF03196391

Caserman, P., Garcia-Agundez, A., Gámez Zerban, A., & Göbel, S. (2021). Cybersickness in current-generation virtual reality head-mounted displays: Systematic review and outlook. *Virtual Reality*, *25*(4), 1153–1170. https://doi.org/10.1007/s10055-021-00513-6

Castellini, O. M., Walejko, G. K., Holladay, C. E., Theim, T. J., Zenner, G. M., & Crone, W. C. (2007). Nanotechnology and the public: Effectively communicating nanoscale science and engineering concepts. *Journal of Nanoparticle Research*, *9*(2), 183–189. https://doi.org/10.1007/s11051-006-9160-z

Castelvecchi, D. (2022). *Black hole at the center of the galaxy imaged for the first time*. *605*, 403–404.

Charles Eames (Director). (1968). *Powers of Ten*.

Chesnutt, K., Gail Jones, M., Corin, E. N., Hite, R., Childers, G., Perez, M. P., Cayton, E., & Ennes, M. (2019). Crosscutting concepts and achievement: Is a sense of size and scale related to achievement in science and mathematics? *Journal of Research in Science Teaching*, *56*(3), 302–321. https://doi.org/10.1002/tea.21511

Chesnutt, K., Jones, M. G., Hite, R., Cayton, E., Ennes, M., Corin, E. N., & Childers, G. (2018). Next generation crosscutting themes: Factors that contribute to students ' understandings of size and scale. *Journal of Research in Science Teaching*, *55*(6), 876–900. https://doi.org/10.1002/tea.21443

Clark, A. (1999). Embodied, situated, and distributed cognition. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 506–517). Blackwell Publishers Ltd.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Lawrence Erlbaum Associates.

Cokelez, A. (2012). Junior High School Students ʾIdeas about the Shape and Size of the Atom. *Research in Science Education*, *42*(4), 673–686. https://doi.org/10.1007/s11165-011-9223-8

College Board. (2009). *College Board Standards for College Success: Science*. College Board.

Collins, A., & Halverson, R. (2009). *Rethinking education in the age of technology: The digital revolution and schooling in America*. Teachers College Press.

Common Core State Standards Initiative. (2023). *Common Core Standards for Mathematics*. Common Core State Standards Initiative.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage Publications.

Corliss, S., & Spitulnik, M. (2008). Student and teacher regulation of learning in technology-enhanced science instruction. *International Perspectives in the Learning Sciences: Cre8ing a Learning World.*, *1*, 167–174.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (Third Edition). SAGE.

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (2023). Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. In M. C. Whitton (Ed.), *Seminal Graphics Papers: Pushing the Boundaries, Volume 2* (1st ed., pp. 51–58). ACM. https://doi.org/10.1145/3596711.3596718

Delgado, C. (2009). *Development of a research-based learning progression for middle school through undergraduate students ʾconceptual understanding of size and scale* [Doctoral Dissertation]. The University of Michigan.

Delgado, C. (2013a). Cross-cultural Study of Understanding of Scale and Measurement: Does the everyday use of US customary units disadvantage US students? *International Journal of Science Education*, *35*(8), 1277–1298. https://doi.org/10.1080/09500693.2013.779761

Delgado, C. (2013b). Navigating Deep Time: Landmarks for Time From the Big Bang to the Present. *Journal of Geoscience Education*, *61*(1), 103–112. https://doi.org/10.5408/12-300.1

Delgado, C., Harper-Gampp, T., Mathenge, R., Peterson, M., & Chen, R. (2023). Impact of VR science lesson on students ʾknowledge of scale. In T. Harper-Gampp, C. Delgado, & R. Mathenge (Eds.), *Leveraging embodied cognition using virtual reality in middle school science education*. [Related paperset]. National Association of Research in Science Teaching 2024, Denver, CO, United States.

Delgado, C., Jones, G., & Parker, D. (2021). Scale, proportion, and quantity. In J. Nordine & O. Lee, *Crosscutting concepts: Strengthening science and engineering learning* (pp. 115–134). National Science Teachers Association.

Delgado, C., & Peterson, M. (2018). An enhanced framework for scale cognition leveraging visual metaphor theory and analogical reasoning theory. In *J. Kay & R. Luckin (Eds.), Rethinking learning in the digital age: Making the learning sciences count (pp. 1607–1608)* (Vol. 3). International Society of the Learning Sciences.

Delgado, C., Stevens, S., Shin, N., Yunker, M., & Krajcik, S. (2007). *The development of students' conceptions of size* [Paper presentation]. National Association for Research in Science Teaching Annual Conference, New Orleans, LA.

Delgado, C., Stevens, S. Y., Shin, N., & Krajcik, J. (2015). A middle school instructional unit for size and scale contextualized in nanotechnology. *Nanotechnology Reviews*, *4*(1). https://doi.org/10.1515/ntrev-2014-0023

Duke Gardens. (2023). *What do plants need?* https://gardens.duke.edu/sites/default/files/What_do_plants_need-2.pdf

Elangovan, N., & Sundaravel, E. (2021). Method of preparing a document for survey instrument validation by experts. *MethodsX*, *8*, 101326. https://doi.org/10.1016/j.mex.2021.101326

Ermi, L., Mäyrä, F., de Castell, S. D., & Jenson, J. (2005). Fundamental components of the gameplay experience: Analyzing immersion. In *Worlds in play: International perspectives on digital games research* (pp. 15–27). Palgrave Macmillan.

Estrada, E., Harper-Gampp, T., Delgado, C., Mathenge, R., Peterson, M., Chen, K. B., & Wu, L. (2023). Co-designing a Science Lesson with VR in Middle School Science. In T. Harper-Gampp, C. Delgado, & R. Mathenge (Eds.), *Leveraging embodied cognition using virtual* *reality in middle school science education*. [Related paperset]. National Association of Research in Science Teaching 2024, Denver, CO, United States.

Ferrell, J. B., Campbell, J. P., McCarthy, D. R., McKay, K. T., Hensinger, M., Srinivasan, R., Zhao, X., Wurthmann, A., Li, J., & Schneebeli, S. T. (2019). Chemical exploration with virtual reality in organic teaching laboratories. *Journal of Chemical Education*, *96*(9), 1961–1966. https://doi.org/10.1021/acs.jchemed.9b00036

Flores, F., Tovar, M. E., & Gallegos, L. (2003). Representation of the cell and its processes in high school students: An integrated view. *International Journal of Science Education*, *25*(2), 269–286. https://doi.org/10.1080/09500690210126793

Fofana, F., Bazeley, P., & Regnault, A. (2020). Applying a mixed methods design to test saturation for qualitative data in health outcomes research. *PLOS ONE*, *15*(6), e0234898. https://doi.org/10.1371/journal.pone.0234898

Franklin, C., Ballan, M., & Thyer, B. A. (2001). Reliability and validity in qualitative research. In *The handbook of social work research methods* (pp. 273–292). Sage Publications.

Gampp, T., Delgado, C., Peterson, M., & Chen, K. B. (2022). Embodied cognition in virtual reality to support learning of scale. In C. Chinn, E. Tan, C. Chan, & Y. Kali (Eds.), *Proceedings of the 16th International Conference of the Learning Sciences—ICLS 2022*. International Society of the Learning Sciences.

Garcia-Pelegrin, E., Wilkins, C., & Clayton, N. S. (2021). The ape that lived to tell the tale. The evolution of the art of storytelling and its relationship to mental time travel and theory of mind. *Frontiers in Psychology*, *12*, 755783. https://doi.org/10.3389/fpsyg.2021.755783

Gerlach, K., Trate, J., Blecking, A., Geissinger, P., & Murphy, K. (2014a). Investigation of absolute and relative scaling conceptions of students in introductory college chemistry courses. *Journal of Chemical Education*, *91*(10), 1526–1537. https://doi.org/10.1021/ed4004707

Gerlach, K., Trate, J., Blecking, A., Geissinger, P., & Murphy, K. (2014b). Valid and reliable assessments to measure scale literacy of students in introductory college chemistry courses. *Journal of Chemical Education*, *91*(10), 1538–1545. https://doi.org/10.1021/ed400471a

Gochman, S. R., Morano Lord, M., Goyal, N., Chow, K., Cooper, B. K., Gray, L. K., Guo, S. X., Hill, K. A., Liao, S. K., Peng, S., Seong, H. J., Wang, A., Yoon, E. K., Zhang, S., Lobel, E., Tregubov, T., & Dominy, N. J. (2019). Tarsier goggles: A virtual reality tool for experiencing the optics of a dark-adapted primate visual system. *Evolution: Education and Outreach*, *12*(1), 9. https://doi.org/10.1186/s12052-019-0101-6

Griffiths, A. K., & Preston, K. R. (1992). Grade-12 students 'misconceptions relating to fundamental characteristics of atoms and molecules. *Journal of Research in Science Teaching*, *29*(6), 611–628. https://doi.org/10.1002/tea.3660290609

Harper-Gampp, T., Delgado, C., Peterson, M., & Chen, K. B. (2023a). *Designing and developing an instrument to assess scale cognition* [Conference presentation]. National Association for Research in Science Teaching 2023, Chicago, IL, United States.

Harper-Gampp, T., Delgado, C., Peterson, M., & Chen, K. B. (2023b). *Refining a panel of experts validation methodology for instrument development* [Roundtable discussion]. American Educational Research Association 2023, Chicago, IL, United States.

Harper-Gampp, T., Delgado, C., Peterson, M., Chen, K. B., Mathenge, R., Planchart, R., & Kulasingam, R. (2023c). Student impressions about a VR science lesson. In T. Harper-Gampp, C. Delgado, & R. Mathenge (Eds.), *Leveraging embodied cognition using virtual reality in middle school science education*. [Related paperset]. National Association of Research in Science Teaching 2024, Denver, CO, United States.

Harper-Gampp, T., Delgado, C., You, H., Peterson, M., & Chen, K. B. (2023d). *A measure of size and scale cognition: The development and validation of the Assessment of Size and Scale Cognition* [Unpublished manuscript].

Hmelo, C. E., Holton, D. L., & Kolodner, J. L. (2000). Designing to Learn About Complex Systems. *Journal of the Learning Sciences*, *9*(3), 247–298. https://doi.org/10.1207/S15327809JLS0903_2

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *Journal of the Learning Sciences*, *16*(3), 307–331. https://doi.org/10.1080/10508400701413401

Hsu, W. C., Tseng, C. M., & Kang, S. C. (2018). Using exaggerated feedback in a virtual reality environment to enhance behavior intention of water-conservation. *Educational Technology and Society*, *21*(4), 187–203.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking: From childhood to adolescence.* Basic Books. https://doi.org/10.1037/10034-000

Integrate. (2018). *Water is essential for food production*. https://serc.carleton.edu/integrate/teaching_materials/food_supply/student_materials/1090#:~:text=Plants need water to grow,soil and into the plant.

Jones, M. G., Chesnutt, K., Ennes, M., Mulvey, K. L., & Cayton, E. (2021). Understanding science career aspirations: Factors predicting future science task value. *Journal of Research in Science Teaching*, *58*(7), 937–955. https://doi.org/10.1002/tea.21687

Jones, M. G., Paechter, M., Yen, C.-F., Gardner, G., Taylor, A., & Tretter, T. (2013). Teachers ' concepts of spatial scale: An international comparison. *International Journal of Science Education*, *35*(14), 2462–2482. https://doi.org/10.1080/09500693.2011.610382

Jones, M. G., Taylor, A., Minogue, J., Broadwell, B., Wiebe, E., & Carter, G. (2007). Understanding scale: Powers of ten. *Journal of Science Education and Technology*, *16*(2), 191–202. https://doi.org/10.1007/s10956-006-9034-2

Jones, M. G., & Taylor, A. R. (2009). Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching*, *46*(4), 460–475. https://doi.org/10.1002/tea.20288

Jones, M. G., Tretter, T., Taylor, A., & Oppewal, T. (2008). Experienced and Novice Teachers ' Concepts of Spatial Scale. *International Journal of Science Education*, *30*(3), 409–429. https://doi.org/10.1080/09500690701416624

Kali, Y., Orion, N., & Eylon, B. (2003). *The effect of knowledge integration activities on students ' perception of the earth 's crust as a cyclic system*. *40*, 545–565.

Kavanagh, S., Luxton-Reilly, A., Wuensche, B., & Plimmer, B. (2017). *A systematic review of Virtual Reality in education*. *10*(2), 85–119.

Khan, F., Pandey, P., & Upadhyay, T. K. (2022). Applications of nanotechnology-based agrochemicals in food security and sustainable agriculture: An overview. *Agriculture*, *12*(10), 1672. https://doi.org/10.3390/agriculture12101672

Kind, V. (2004). *Beyond Appearances: Students ' misconceptions about basic chemical ideas*. Report prepared for the Royal Society of Chemistry, London.

Kong, Y., Douglas, K. A., Rodgers, K. J., Diefes-Dux, H., & Madhavan, K. (2017). A size and scale framework for guiding curriculum design and assessment. *Journal of Engineering Education*, *106*(3), 431–453. eric. https://doi.org/10.1002/jee.20172

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lancaster, S. J. (2018). Immersed in virtual molecules. *Nature Reviews Chemistry*, *2*(10), 253–254. https://doi.org/10.1038/s41570-018-0043-5

Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., Fata-Hartley, C. L., Ebert-May, D., Jardeleza, S. E., & Cooper, M. M. (2016). Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLOS ONE*, *11*(9), e0162333. https://doi.org/10.1371/journal.pone.0162333

Lee, K. M. (2004). Presence, explicated. *Communication Theory*, *14*(1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lee, T. D., Gail Jones, M., & Chesnutt, K. (2019). Teaching Systems Thinking in the Context of the Water Cycle. *Research in Science Education*, *49*(1), 137–172. https://doi.org/10.1007/s11165-017-9613-7

Light, G., Swarat, S., Park, E. J., Drane, D., Tevaarwerk, E., & Mason, T. (2007). Understanding undergraduate students ' conceptions of a core nanoscience concept: Size and scale. *Proceedings of the Inaugural International Conference on Research in Engineering Education, ICREE*.

Lin, S.-F., Lin, H., & Wu, Y. (2013). Validation and exploration of instruments for assessing public knowledge of and attitudes toward nanotechnology. *Journal of Science Education and Technology*, *22*(4), 548–559. https://doi.org/10.1007/s10956-012-9413-9

Linn, M. C., & Eylon, B.-S. (2011). *Science Learning and Instruction*. Routledge. https://doi.org/10.4324/9780203806524

Liu, R., Wang, L., Lei, J., Wang, Q., & Ren, Y. (2020). Effects of an immersive virtual reality-based classroom on students ' learning performance in science lessons. *British Journal of Educational Technology*, *51*(6), 2034–2049. https://doi.org/10.1111/bjet.13028

Loftus, E. F., & Marburger, W. (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition*, *11*(2), 114–120. https://doi.org/10.3758/BF03213465

Lombard, M., & Ditton, T. (2006). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2), 0–0. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Luo, H., Li, G., Feng, Q., Yang, Y., & Zuo, M. (2021). Virtual reality in K-12 and higher education: A systematic review of the literature from 2000 to 2019. *Journal of Computer Assisted Learning*, *37*(3), 887–901. https://doi.org/10.1111/jcal.12538

Madhok, J. (2006). *The longitudinal impact of an eighth-grade inquiry curriculum on students' beliefs and achievement in science* [Doctoral Dissertation, University of California, Berkeley]. https://www.proquest.com/dissertations-theses/longitudinal-impact-eighth-grade-inquiry/docview/305348548/se-2

Magaña, A. J. (2014). Learning strategies and multimedia techniques for scaffolding size and scale cognition. *Computers & Education*, *72*, 367–377. https://doi.org/10.1016/j.compedu.2013.11.012

Magana, A. J., Brophy, S. P., & Bryan, L. A. (2012a). An integrated knowledge framework to characterize and scaffold size and scale cognition (FS2C). *International Journal of Science Education*, *34*(14), 2181–2203. https://doi.org/10.1080/09500693.2012.715316

Magaña, A., Newby, T., & Brophy, S. (2012b). *Comparing novice and expert perceptions of interactive multimedia tools for conveying conceptions of size and scale*. *20*(4), 441–465.

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*, *33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Maksimenko, N., Okolzina, A., Vlasova, A., Tracey, C., & Kurushkin, M. (2021). Introducing atomic structure to first-year undergraduate chemistry students with an immersive virtual reality experience. *Journal of Chemical Education*, *98*(6), 2104–2108. https://doi.org/10.1021/acs.jchemed.0c01441

Mandrikas, A., Michailidi, E., & Stavrou, D. (2020). Teaching Nanotechnology in Primary Education. *Research in Science & Technological Education*, *38*(4), 377–395. https://doi.org/10.1080/02635143.2019.1631783

Mathenge, R., Kulasingam, R., Delgado, C., Peterson, M., & Chen, K. B. (2023). Impact of an VR Science Lesson on Reform-Oriented Nature of Science Instruction. In T. Harper-Gampp, C. Delgado, & R. Mathenge (Eds.), *Leveraging embodied cognition using virtual reality in middle school science education*. [Related paperset]. National Association of Research in Science Teaching 2024, Denver, CO, United States.

Matovu, H., Ungu, D. A. K., Won, M., Tsai, C.-C., Treagust, D. F., Mocerino, M., & Tasker, R. (2023). Immersive virtual reality for science learning: Design, implementation, and evaluation. *Studies in Science Education*, *59*(2), 205–244. https://doi.org/10.1080/03057267.2022.2082680

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students 'learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, *70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

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 & Education*, *140*, 103603. https://doi.org/10.1016/j.compedu.2019.103603

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, *56*(3), 769–780. https://doi.org/10.1016/j.compedu.2010.10.020

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, *46*(6), 675–698. https://doi.org/10.1002/tea.20314

National Academies of Sciences, Engineering, and Medicine. (2020). *A quadrennial review of the National Nanotechnology Initiative: Nanosciences, applications, and commercialization.* The National Academies Press.

National Aeronautics and Space Administration. (2022). *NASA 2022 strategic plan*. National Aeronautics and Space Administration.

National Air and Space Museum. (1986). *IMAX® Audiences Embark on a Cosmic Voyage Through Time and Space*. National Air and Space Museum.

National Nanotechnology Initiative. (2021). *National nanotechnology initiative strategic plan*. Executive Office of the President of the United States, National Science and Technology Council.

National Research Council. (1996). *National science education standards*. National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.

National Science Foundation. (2020). *STEM education for the future: A visioning report*. National Science Foundation.

Network for Computational Nanotechnology. (2007). *Generation nano* [Computer software]. http://www.generation-nano.org/

Next Generation Science Standards Lead States. (2013). *Next Generation Science Standards*. The National Academies Press.

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*, 152–158. https://doi.org/10.1145/191666.191729

Nikon. (2009). *Universcale* [Computer software]. http://www.nikon.com/about/feelnikon/universcale/index.html

Nordine, J., & Lee, O. (2021). *Crosscutting concepts: Strengthening science and engineering learning*. National Science Teaching Association.

North Carolina Department of Public Instruction. (2024). *North Carolina Essential Science Standards: 6-8 science*. North Carolina Department of Public Instruction.

Olson, S., & Loucks-Horsley, S. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. National Academies Press.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Paivio, A. (1990). *Mental representations: A dual coding approach*. Clarendon Press.

Parkline Interactive. (2019). *Wander* [Computer software]. https://www.oculus.com/deeplink/?action=view&path=/app/2078376005587859&ref=oculus_desktop

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

Playgen. (2009). *NanoMission: Learning scale from pico to giga* [Computer software]. http://nanomission.org/category/nanoscaling/

Resnick, I., Davatzes, A., Newcombe, N., & Shipley, T. (2017). Using relational reasoning to learn about scientific phenomena at unfamiliar scales. *EDUCATIONAL PSYCHOLOGY REVIEW*, *29*(1), 11–25. https://doi.org/10.1007/s10648-016-9371-5

Ryoo, K., & Linn, M. C. (2012). Can dynamic visualizations improve middle school students ' understanding of energy in photosynthesis? *Journal of Research in Science Teaching*, *49*(2), 218–243. https://doi.org/10.1002/tea.21003

Sabelli, N., Schank, P., Rosenquist, A., Stanford, T., Cormia, R., & Hurst, K. (2005). *Report of the workshop science and technology education at the nanoscale*. SRI International. nanosense.org/documents/reports/NanoWorkshopReportDraft.pdf

Sadler, T. D. (2011). *Socio-scientific Issues in the Classroom* (Vol. 39). Springer. https://doi.org/10.1007/978-94-007-1159-4

Sakhnini, S., & Blonder, R. (2015). Essential concepts of nanoscale science and technology for high school students based on a delphi study by the expert community. *International*

Journal of Science Education, *37*(11), 1699–1738. https://doi.org/10.1080/09500693.2015.1035687

Singh, P. M., Tiwari, A., Maity, D., & Saha, S. (2022). Recent progress of nanomaterials in sustainable agricultural applications. *Journal of Materials Science*, *57*(24), 10836–10862. https://doi.org/10.1007/s10853-022-07259-9

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

Steffen, J. H., Gaskin, J. E., Meservy, T. O., Jenkins, J. L., & Wolman, I. (2019). Framework of affordances for virtual reality and augmented reality. *Journal of Management Information Systems*, *36*(3), 683–729. https://doi.org/10.1080/07421222.2019.1628877

Swarat, S., Light, G., Park, E. J., & Drane, D. (2011). A typology of undergraduate students ’ conceptions of size and scale: Identifying and characterizing conceptual variation. *Journal of Research in Science Teaching*, *48*(5), 512–533. https://doi.org/10.1002/tea.20403

Tilhou, R., Taylor, V., & Crompton, H. (2020). 3D Virtual Reality in K-12 Education: A Thematic Systematic Review. In S. Yu, M. Ally, & A. Tsinakos (Eds.), *Emerging Technologies and Pedagogies in the Curriculum* (pp. 169–184). Springer Singapore. https://doi.org/10.1007/978-981-15-0618-5_10

Trate, J. M., Hackl, A., Mohs, B., Heinze, K., Geissinger, P., Blecking, A., & Murphy, K. L. (2019). Classwide investigation of absolute and relative scaling conceptions of students in introductory college chemistry. *Journal of Chemical Education*, *96*(7), 1341–1350. eric. https://doi.org/10.1021/acs.jchemed.8b00988

Trend, R. D. (2001). Deep time framework: A preliminary study of U.K. primary teachers ’ conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching*, *38*, 191–221.

Tretter, T. R., Jones, M. G., Andre, T., Negishi, A., & Minogue, J. (2006a). Conceptual boundaries and distances: Students ’and experts ’concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching*, *43*(3), 282–319. https://doi.org/10.1002/tea.20123

Tretter, T. R., Jones, M. G., & Minogue, J. (2006b). Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching*, *43*(10), 1061–1085. https://doi.org/10.1002/tea.20155

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. *Spatial Information Theory: A Theoretical Basis for GIS*, 14–24.

Vergnaud, G. (1988). Multiplicative structures. In J. Hiebert & M. Behr, *Number Concepts and Operation in the Middle Grades* (pp. 141–161).

Vlaardingerbroek, B., Taylor, N., & Bale, C. (2014). The problem of scale in the interpretation of pictorial representations of cell structure. *Journal of Biological Education*, *48*(3), 154–162. eric. https://doi.org/10.1080/00219266.2013.849284

Waldron, A. M., Spencer, D., & Batt, C. A. (2006). The current state of public understanding of nanotechnology. *Journal of Nanoparticle Research*, *8*(5), 569–575. https://doi.org/10.1007/s11051-006-9112-7

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Winkelmann, K., & Bhushan, B. (Eds.). (2016). *Global Perspectives of Nanoscience and Engineering Education*. Springer International Publishing. https://doi.org/10.1007/978-3-319-31833-2

Wu, L., Sekelsky, B., Peterson, M., Gampp, T., Delgado, C., & Chen, K. B. (2022). Immersive virtual environment for scale cognition and learning: Expert-based evaluation for balancing usability versus cognitive theories. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1972–1976. https://doi.org/10.1177/1071181322661094

Xie, Y., Hunter, M., Sorensen, A., Nogeire-McRae, T., Murphy, R., Suraci, J. P., Lischka, S., & Lark, T. J. (2023). U.S. farmland under threat of urbanization: Future development scenarios to 2040. *Land*, *12*(3), 574. https://doi.org/10.3390/land12030574

Youngblut, C. (1998). *Educational uses of virtual reality technology*. Institute of Defense Analyses (IDA Document D-2128).

APPENDICES

Appendix A: Interview Protocol

Introduction

Thank you for agreeing to participate in this interview. We are interviewing you to better understand how virtual reality can be used to support students' understanding of the size of different objects. Therefore, there are no right or wrong answers to any of our questions, we are interested in your own experience.

This interview should take approximately 15 minutes depending on how much information you would like to share. With your permission, I would like to audio record the interview because I don't want to miss any of your comments. You may stop the interview at any time and for any reason. Are there any questions about what I have just explained?

May I begin the audio recording?

Please note that this guide only represents the main themes to be discussed with the students and as such does not include the various prompts that may also be used. Non-leading and general prompts will also be used, such as "Can you please tell me a little bit more about that?" and "What does that look like?"

Establish Rapport

Before we begin, it would be nice if you could tell me a little bit about yourself. [Tailor a question here to a specific person and/or situation. For example: " You have a beautiful school, can you tell me what is your favorite part?"]

1. **Whole-class Activity**
   How would you describe your experience in virtual reality (Scale Worlds & Wonder)/using Universcale during our class activities?
      Prompt: What particular parts, objects, or things stood out to you in Scale Worlds/Universcale? What did you like? What did you not like? What did you learn? What helped you learn? What parts of the activity did you enjoy? How would you describe the participation of your peers during the activity? How do you think virtual reality can be used in class to help you learn? Was there anything in particular that you did not like about the activity?

2. **Assessment of Size and Scale Cognition [Survey]**
   On the survey, you were asked to order a set of entities (*include handout with student responses).* Can you tell me a little bit about what caused the change in your responses?

On the survey, you were asked to identify the absolute size of various objects or determine which object was a particular size. Can you describe the way you approached these problems? (***Prompt****: Why might your responses change?*)

Prompt: How did the Scale World's experience change the way you ordered the objects? How did Scale Worlds impact the way you approached the assessment? Were there parts of the assessment that you felt more confident on after experiencing Scale Worlds?
Prompt: Tell me why you ordered the following objects as cell, amoeba, molecule.

3. Disciplinary Core Ideas

How do humans impact cropland?

How does water help crops grow and survive?

How are the water molecules in a plant cell connected to the water molecules in the Earth's oceans?

What part of the activities do you feel helped you to understand these concepts?

4. Conclusion

Would you recommend that your teachers use virtual reality/Universcale to help support your learning?

Prompt: Can you elaborate on ways it might be used in other subjects? Is there anything else that you would like to comment on about Scale Worlds, or the activity that we haven't discussed today? Would you recommend that your teachers use Scale Worlds? Would you recommend they use either desktop or head-mounted display, and why?

Appendix B: Feedback Survey

1. **Likert Scale Statements** *(5-point scale: Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree)***:**
   A. I enjoyed using Scale Worlds VR.
   B. I learned something while experiencing Scale Worlds VR.
   C. I think that using virtual reality is an effect way to learn about the size of various objects.
   D. Scale Worlds VR increased my understanding of the size of various objects.
2. **Free Response Items:**
   A. Describe something you learned about the size of one object while using Scale Worlds VR.
   B. What was your favorite thing about Scale Worlds VR?
   C. What did you not like about Scale Worlds VR?
   D. How would you feel about learning with virtual reality in your other classes? Please describe.
   E. Have you ever wanted to use virtual reality headsets for something other than classroom learning and gaming? Please describe.
   F. Tell me about a situation where using virtual reality would be really useful.
   G. For the situation you described in the previous question, are there any obstacles for you to use VR in that situation (assuming you had a VR headset available)?
3. **Multiple Choice**:
   A. Did you experience the following feelings while using Scale Worlds? [Yes/No]
      a. Awe (amazement, astonishment, or wonder)
         i. If student indicated yes, they were asked the following follow up question: what caused those feelings of awe (amazement, astonishment, or wonder)? Please describe.
      b. Excitement (amusement, joy, or delight)
      c. Frustration (annoyance or aggravation)
      d. Fear (anxiety, nervousness, or worry)

i. If student indicated yes, they were asked the following follow up question: what caused this feeling of fear (anxiety, nervousness, or worry)? Please describe.

B. Scale Worlds VR was made by STEM researchers to help prepare students for STEM careers. STEM includes science, technology and design, engineering, and math. Did Scale Worlds VR change your interest in working in STEM in the future?

   a. I am more interested in working in STEM in the future.
   b. No change in my interest in working in STEM in the future.
   c. I am less interested in working in STEM in the future.

Appendix C: Expanded DCI Codebook

DCI	Description	Rules	*Example(s)*
LS1.A Structure and Function	"They have characteristic structures (anatomy and morphology), functions (molecular-scale processes to organism-level physiology), and behaviors (neurobiology and, for some animal species, psychology)." (NRC, 2012, p. 143)	Focuses on the structure, function, and behaviors (*water giving nutrients is NOT a function but water transporting nutrients through out the plant is*). Another example might be roots absorbing water (a function of the roots).	*(1) Plants Use Water to **Transport Nutrients and Minerals**; (2) water makes the **roots like stand up**, hold on to themselves.*
LS1.C Organization for Matter and Energy Flow in Organisms	"Energy needed for life is ultimately derived from the sun through photosynthesis. Plants, algae (including phytoplankton), and other energy fixing microorganisms use sunlight, water, and carbon dioxide to facilitate photosynthesis…plants and algae - being the resource base for animals…Plants need water and light to live and grow." (NRC, 2012, p. 147)	Must indicate a NEED for water not just hydration or moisture (e.g., water or it will die); Any reference to photosynthesis is included.	*(1) They **need water** in order to survive.; (2) It supplies them with **oxygen and hydrogen,** that will then get **turned into glucose;** (3) it **gives them nutrients***
ESS2.C The Roles of Water in Earth's Surface Processes	"Water is found almost everywhere on Earth, from high in the atmosphere (as water vapor and ice crystals) to low in the atmosphere (precipitation, droplets in clouds) to mountain snowcaps and glaciers (solid) to running liquid water on the land, ocean, and underground. Energy from the sun and the force of gravity drive the continual cycling of water among these reservoirs." (NRC, 2012, p. 184)	Response describes either a logical place to find water OR the cycling of water. (i.e., getting water from the roots is a function of the plant (LS1.A); saying that the roots absorb water stored in the ground (LS1.A & ESS2.C))	*(1) Through **in the ground** because water flows everywhere; (2) Crops can **get water naturally from rain** or get watered by humans;*

Expanded DCI Codebook (continued).

ESS3.C Humans Impacts on Earth Systems	"Human activities in agriculture, industry, and everyday life have had a major impact on the land, rivers, ocean, and air…and because all of Earth's subsystems are interconnected, changes in one system can produce unforeseen changes in others." (NRC, 2012, p. 195) This includes the "distribution of Earth's water through the modification of streams, lakes, and ground- water (NRC, 2012, p. 194)." "Some negative effects of human activities are reversible with informed and responsible management. For example, communities are doing many things to help protect Earth's resources and environments. They are treating sewage, reducing the amount of materials they use, and reusing and recycling materials." (NRS, 2012, p. 195)	Must describe an impact to the biosphere (e.g., "a decrease in food" does not indicate a change to the biosphere it is an impact on people). This included human developed watering systems (e.g., sprinklers, by hand, machines) which must divert water. For items 6 & 7, coded if item listed is a reasonable human pollutant (e.g., trash, chemicals, plastics).	*(1) Crops can get water naturally from rain or **get watered by humans;***
PS1.A Structure and Propertzies of Matter	"Pure substances are made from a single type of atom or molecule; each pure substance has characteristic physical and chemical properties (for any bulk quantity under given conditions) that can be used to identify it." (NRC, 2012, p. 108)	For item 5, ESS2.C unless student specifically identifies the molecular structure	*(1) its both h20 the **same element***

Appendix D: Expanded SPQ Codebook

SPQ	Description	Rules	Example(s)
Scale	(1) the comparison of objects, processes, or properties across size, time, or energy scale, or to dimensions of familiar objects, timescale, or energies or (2) the identification of non-negligible/relevant interactions at more than one scale (Laverty et al., 2016, p. 8)	Objects, processes, or properties must vary in size accordingly: microscopic scale (<10-3 m), human scale (103 m), and big scale (>103 m) (based on findings from Tretter et al., 2006CON) For human activities, the activities themselves were considered at human scale (e.g., watering, planting, harvesting, growing, building homes) unless scale was otherwise indicated. For example, the activity of genetically modifying requires manipulations at the microscopic scale as thus was considered microscopic not human while planting was considered human scale. "The population" is assumed to mean human population. For item 4, 9, 10, human scale is implied in the question ("a crop" & "a farmer"). Therefore, responses were coded for reference to either microscopic or big.	*(1) Human activities make cropland decrease. Some ways they can do this is by* ***contaminating the soil (**<u>human scale</u>**)*** *so crops can't grow,* ***polluting the land (**<u>human scale</u>**)*** *and water, and* ***urbanizing areas (**<u>big scale</u>**)*** *where cropland was for humans to have houses and resources that they want/need that are not crops.; (2) People pollute (<u>human scale)</u> the water and put really bad chemicals (<u>human scale</u>) in some of there things to kill the bugs (<u>ambiguous</u>). [Not coded]*

Expanded SPQ Codebook (continued).

Proportion & Quantity	1) Identifying a response of one quantitative variable to changes in another or (2) the relationship between two or more quantitative variables (Laverty et al., 2016, p. 8) "Grasp the relationships among different types of quantities" which is "different than a ratio of numbers describing fractions of a pie. (NRC, 2012, p. 90)"	The variables must be quantitative (e.g., increases, decreases, more, less) not categorical (e.g., it is present or not). Not "helps" growth but that it "increases" or makes "faster" or "expands."	*The population is **growing**, meaning that there will **be less space for farming** and farmers will have to **grow crops at a much faster rate** in order to feed those in need.*

Appendix E: Assessing Collective Landmarks

F.1 Pretest absolute size values for VR participating students from the 5th through 95th percentiles.

Entity	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	M	STD
Sun (Diameter)	-1	2	2	2	3	4	6	8	**9**	**9**	**9**	**9**	**9**	**9**	**9**	10	10	10	10	6.1	5.3
Saturn (Diameter)	-1	2	2	3	3	3	4	5	7	**8**	**8**	9	9	9	10	10	10	10	11	5.9	5.0
Venus (Diameter)	0	1	1	2	2	2	3	4	4	4	4	5	**7**	8	8	8	8	8	8	4.4	3.5
Mercury (Diameter)	-1	1	2	3	4	5	**6**	**6**	**6**	**6**	7	7	7	8	9	9	10	10	11	5.9	3.8
Panama Canal (Length)	-5	-2	-1	-0	2	2	3	3	3	3	3	4	4	4	4	4	**5**	**5**	7	2.2	4.1
Mount Everest (Height)	-1	3	**4**	**4**	**4**	5	6	6	6	6	7	7	8	9	10	11	11	11	11	6.2	4.6
Golden Gate Bridge (Length)	-2	-0	0	0	1	1	1	1	2	2	**3**	**3**	**3**	4	5	5	6	6	8	2.7	3.5
Titanic (Length)	-1	1	1	1	**2**	3	3	3	3	5	5	5	5	6	6	6	7	8	9	4.0	3.7
School Bus (Length)	-7	-4	-2	-2	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	2	3	4	5	6	0.6	4.0
Human (Height)	-5	-4	-4	-3	-3	-2	-1	-1	-1	-1	**-0**	**0**	**0**	**0**	**0**	**0**	**0**	2	5	-1.0	3.0
Squirrel (Height)	-10	-10	-9	-7	-5	-4	-3	-3	-3	-2	-2	**-1**	**-1**	**-1**	**-1**	-0	1	3	6	-2.8	5.0
Bumblebee (Length)	-11	-10	-10	-9	-9	-7	-6	-5	-3	**-2**	**-2**	**-2**	**-2**	**-2**	-1	-1	-1	0	3	-4.1	4.7
Fruit Fly (Length)	-12	-12	-11	-9	-8	-6	-5	-5	-5	-4	**-3**	**-3**	**-3**	-2	-2	-2	-1	-1	1	-4.9	4.4
Human Egg Cell (Diameter)	-11	-11	-11	-11	-10	-9	-8	-7	-7	-5	-5	-5	-5	**-4**	**-4**	**-4**	-3	-3	-2	-6.3	3.9
Skin Cell (Diameter)	-8	-8	-8	-8	-6	**-5**	**-5**	**-5**	**-5**	**-5**	**-5**	**-5**	**-5**	-4	-3	-3	-2	1	6	-3.9	4.6
Nucleus of a Cell (Diameter)	-11	-11	-11	-10	-10	-8	-7	**-6**	**-6**	**-6**	-5	-4	-4	-4	-4	-2	-2	-0	2	-5.6	4.4
Influenza Virus (Diameter)	-11	-11	-11	-10	-8	-8	**-7**	**-7**	-6	-5	-5	-5	-5	-4	-3	-2	-1	1	4	-5.2	5.1
Ribosome (Length)	-10	-10	-10	-10	-9	-9	-9	-9	**-8**	-7	-7	-7	-7	-6	-5	-2	0	2	4	-5.9	5.1
Glucose Molecule (Length)	-11	-10	-10	-10	-10	-10	**-9**	**-9**	-8	-8	-6	-6	-6	-5	-3	1	2	2	4	-5.5	5.7
Carbon Dioxide Molecule (Length)	**-10**	-9	-9	-8	-8	-7	-7	-7	-7	-6	-6	-5	-5	-4	-4	-3	2	6	8	-4.3	6.0

F.2 Posttest absolute size values for VR participating students from the 5th through 95th percentiles.

Entity	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	M	STD
Sun (Diameter)	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	9	9	10	11	9.2	0.7
Saturn (Diameter)	-8	-1	3	4	4	5	5	5	6	6	6	7	7	7	7	**8**	**8**	**8**	9	4.5	5.4
Venus (Diameter)	-1	1	2	2	4	4	4	5	6	**7**	**7**	**7**	**7**	**7**	8	8	8	8	8	4.9	3.8
Mercury (Diameter)	1	2	2	3	3	4	5	5	5	5	**6**	**6**	**6**	**6**	7	7	7	7	8	5.0	2.7
Panama Canal (Length)	-11	-10	-8	-6	-4	-1	1	1	2	4	4	4	4	4	4	6	6	6	7	0.5	6.4
Mount Everest (Height)	-2	1	2	2	3	3	**4**	**4**	**4**	**4**	**4**	**4**	**4**	5	5	6	6	6	6	3.5	2.7
Golden Gate Bridge (Length)	2	2	2	2	**3**	**3**	**3**	**3**	**3**	4	4	4	4	5	7	8	8	9	11	4.8	3.1
Titanic (Length)	-1	-1	-1	0	1	1	1	1	**2**	**2**	3	3	3	3	5	5	6	7	9	2.9	2.9
School Bus (Length)	-8	-7	-5	-2	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	2	3	4	-0.4	3.7
Human (Height)	-7	-7	-5	-3	-3	-2	-1	-1	-1	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	1	2	-1.6	3.3
Squirrel (Height)	-8	-5	-4	-4	-3	-3	-2	-2	-2	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	-0	1	2	5	-1.5	4.4
Bumblebee (Length)	-9	-8	-8	-7	-6	-5	-4	-3	-3	-3	-1	-1	-1	-1	-1	-1	-0	1	4	-2.8	5.0
Fruit Fly (Length)	-8	-7	-6	-5	-4	**-3**	**-3**	**-3**	-2	-2	-2	-2	-2	-2	-2	-2	-1	-0	2	-2.9	3.4
Human Egg Cell (Diameter)	-10	-10	-9	-8	-7	-5	-5	-5	-5	-5	**-4**	**-4**	**-4**	**-4**	-3	-2	-1	1	5	-4.3	4.4
Skin Cell (Diameter)	-9	-8	-8	-8	-7	-6	-6	-6	-6	-4	-3	-3	-3	-3	-2	-0	1	2	4	-3.7	4.7
Nucleus of a Cell (Diameter)	-11	-10	-10	-9	-9	-8	-8	-8	-8	-7	**-6**	-5	-5	-4	-3	0	3	5	5	-4.8	5.9
Influenza Virus (Diameter)	-11	-10	-10	-10	-9	-9	-8	**-7**	**-7**	**-7**	**-7**	**-7**	**-7**	-6	-6	-4	-0	2	2	-6.2	4.3
Ribosome (Length)	-11	-11	-11	-11	-9	**-8**	**-8**	**-8**	**-8**	**-8**	-6	-5	-5	-5	-4	-2	-1	1	5	-5.8	5.1
Glucose Molecule (Length)	-10	-10	-10	**-9**	-8	-7	-7	-7	-7	-6	-6	-6	-6	-6	-4	-3	-1	1	4	-5.2	5.4
Carbon Dioxide Molecule (Length)	-12	**-10**	-9	-8	-8	-8	-7	-6	-6	-6	-6	-5	-5	-5	-4	-2	-1	0	2	-5.4	4.3

F.3 Pretest absolute size values for non-VR participating students from the 5th through 95th percentiles.

Entity	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	M	STD
Sun (Diameter)	-4	1	6	6	7	7	7	8	8	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	10	10	6.1	6.7
Saturn (Diameter)	2	2	3	4	5	5	6	7	**8**	9	10	10	11	11	11	11	11	11	11	7.5	4.0
Venus (Diameter)	-6	-2	2	2	2	2	2	3	3	5	6	6	6	6	6	**7**	**7**	**7**	8	3.0	5.7
Mercury (Diameter)	3	3	4	4	4	4	**6**	7	8	8	8	9	10	11	11	11	11	11	11	7.4	3.6
Panama Canal (Length)	-4	-4	-3	-3	-2	-2	-1	0	1	2	3	3	4	4	4	**5**	**5**	**5**	**5**	1.0	3.9
Mount Everest (Height)	-8	-7	-6	-4	-2	0	0	1	1	3	**4**	**4**	**4**	**4**	5	5	6	7	9	1.3	6.3
Golden Gate Bridge (Length)	**3**	**3**	4	4	4	4	4	4	4	4	4	5	6	7	8	8	9	10	10	5.6	3.1
Titanic (Length)	-2	-2	-2	-1	0	1	1	**2**	**2**	3	3	3	3	3	3	3	3	3	4	1.5	2.3
School Bus (Length)	-7	-7	-6	-5	-5	-4	-2	-1	0	0	0	0	0	0	0	**1**	**1**	**1**	**1**	-2.0	3.5
Human (Height)	-9	-9	-9	-9	-8	-8	-6	-4	-3	-3	-2	-2	-2	-2	-1	**0**	2	3	5	-3.1	5.4
Squirrel (Height)	-9	-9	-8	-7	-7	-6	-6	-5	-4	-3	-2	**-1**	-0	-0	0	1	1	2	4	-3.0	5.1
Bumblebee (Length)	-12	-11	-11	-10	-10	-9	-8	-7	-7	-6	-5	-5	-5	-5	-5	-4	-3	-1	2	-6.0	5.1
Fruit Fly (Length)	-11	-11	-10	-10	-9	-9	-9	-9	-9	-9	-8	-7	-6	-4	**-3**	-2	-1	-1	-1	-6.8	4.2
Human Egg Cell (Diameter)	-11	-11	-10	-10	-10	-9	-7	-5	**-4**	-3	-2	-2	-2	-2	-1	-0	1	1	1	-4.8	5.2
Skin Cell (Diameter)	-10	-9	-9	-9	-8	-8	-7	-6	**-5**	**-5**	**-5**	-4	-3	-1	0	2	4	6	7	-3.1	6.7
Nucleus of a Cell (Diameter)	-12	-12	-12	-12	-12	-11	-9	-7	**-6**	-5	-4	-3	-1	0	1	2	2	2	3	-5.0	6.5
Influenza Virus (Diameter)	-12	-12	-12	-12	-12	-11	-8	-6	-4	-4	-3	-3	-3	-3	-3	-2	-2	0	3	-5.4	6.1
Ribosome (Length)	-9	**-8**	**-8**	-7	-7	-6	-4	-2	-1	-1	-1	-0	1	2	2	3	3	4	6	-1.6	5.7
Glucose Molecule (Length)	-12	-12	-12	-12	-11	-11	**-9**	-7	-6	-6	-6	-5	-4	-3	-3	-2	-1	-0	1	-6.1	5.3
Carbon Dioxide Molecule (Length)	-12	-12	-12	-12	-12	-12	**-10**	-8	-7	-6	-5	-3	1	4	5	5	5	6	8	-3.6	8.7

F.4 Posttest absolute size values for non-VR participating students from the 5th through 95th percentiles.

Entity	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	M	STD
Sun (Diameter)	6	7	7	7	7	7	8	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	8.1	
Saturn (Diameter)	3	3	4	5	6	7	7	**8**	**8**	9	9	9	9	9	9	10	10	10	10	7.4	
Venus (Diameter)	1	2	2	3	4	4	4	4	4	4	4	4	5	5	6	6	**7**	**7**	8	4.4	
Mercury (Diameter)	1	1	2	2	3	3	3	3	3	4	4	4	5	5	5	5	5	**6**	7	3.8	
Panama Canal (Length)	-2	-2	-1	0	1	2	2	2	2	3	3	3	3	3	3	3	3	4	6	2.0	
Mount Everest (Height)	-1	0	1	1	2	2	2	2	2	2	2	3	**4**	5	5	6	6	7	8	3.1	
Golden Gate Bridge (Length)	-4	-3	-3	-0	2	4	4	4	4	4	4	4	5	5	6	7	8	8	8	3.3	
Titanic (Length)	-6	-3	-1	-0	1	1	**2**	3	3	3	3	3	4	5	6	6	7	8	8	2.4	
School Bus (Length)	-2	-1	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	2	2	4	6	1.3	
Human (Height)	-5	-4	-2	-1	-1	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	1	2	4	5	6	0.3	
Squirrel (Height)	-6	-5	-5	-5	-5	-5	-4	-3	-2	-2	-2	-2	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	-2.9	
Bumblebee (Length)	-4	-3	-3	-3	-3	-3	-3	**-2**	**-2**	**-2**	**-2**	**-2**	-1	-1	-1	0	1	3	6	-0.8	
Fruit Fly (Length)	-9	-8	-6	-6	-6	-6	-4	**-3**	-2	-2	-2	-2	-2	-2	-2	-1	-1	-0	0	-3.6	
Human Egg Cell (Diameter)	-9	-8	-8	-8	-7	-7	-6	-5	-5	**-4**	-3	-3	-2	-2	-2	-1	-1	-0	0	-4.3	
Skin Cell (Diameter)	-6	-6	**-5**	**-5**	**-5**	**-5**	**-5**	-4	-4	-4	-3	-3	-3	-3	-3	-2	-1	-0	0	-3.4	
Nucleus of a Cell (Diameter)	-10	-9	-8	-8	-8	-8	-7	**-6**	-5	-5	-4	-2	1	4	6	6	7	7	8	-2.0	
Influenza Virus (Diameter)	-10	-10	-10	-10	-9	-9	-8	**-7**	**-7**	-6	-4	-4	-3	-2	-2	-1	-0	2	4	-4.5	
Ribosome (Length)	-10	-10	-10	-10	-9	-9	**-8**	-7	-6	-5	-4	-3	-2	-1	0	2	4	5	5	-3.8	
Glucose Molecule (Length)	-11	-10	-10	-10	-10	-10	-10	**-9**	-8	-7	-6	-5	-4	-4	-2	0	3	3	4	-5.3	
Carbon Dioxide Molecule (Length)	-12	-12	-12	-12	-11	-11	-11	**-10**	**-10**	**-10**	-9	-9	-8	-7	-7	-6	-5	-3	-0	-8.0	

F.5 Pretest relative size values for VR participating students from the 5th through 95th percentiles.

Entity	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	M	STD
Sun (Diameter)	2	4	5	6	6	7	7	**9**	**9**	**9**	**9**	**9**	**9**	**9**	10	10	11	14	16	8.5	
Saturn (Diameter)	-3	-3	-1	0	2	2	2	3	5	**8**	**8**	**8**	**8**	10	13	14	14	16	17	6.5	
Venus (Diameter)	-1	0	2	2	3	4	4	4	4	4	4	5	**7**	8	8	8	8	9	9	4.7	
Mercury (Diameter)	0	1	1	2	3	4	4	4	4	5	**6**	**6**	7	7	7	8	9	9	11	5.3	
Panama Canal (Length)	-3	-2	0	1	2	3	3	3	3	3	3	3	3	3	4	**5**	6	7	11	3.4	
Mount Everest (Height)	1	1	1	1	2	3	**4**	6	6	6	6	6	8	9	10	11	11	11	11	6.0	
Golden Gate Bridge (Length)	-2	0	2	2	2	2	2	**3**	**3**	4	4	4	5	5	6	7	7	9	11	4.1	
Titanic (Length)	-3	-2	-1	0	1	1	1	**2**	**2**	**2**	3	4	5	5	5	5	6	7	8	2.7	
School Bus (Length)	-3	-1	-1	-1	-1	-0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	1	1	3	5	0.6	
Human (Height)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Squirrel (Height)	-6	-5	-4	-4	-3	-3	-3	-2	-2	-2	-2	-2	**-1**	**-1**	**-1**	**-1**	**-1**	-0	2	-1.9	
Bumblebee (Length)	-5	-5	-4	-4	-4	-3	-3	-3	-3	-3	**-2**	**-2**	**-2**	**-2**	**-2**	-1	-1	-1	1	-2.3	
Fruit Fly (Length)	-12	-9	-7	-6	-5	-5	-5	-5	-5	-5	-4	-4	**-3**	**-3**	**-3**	-2	-2	-1	-0	-4.8	
Human Egg Cell (Diameter)	-14	-12	-11	-9	-8	-7	-7	-6	-6	-5	**-4**	**-4**	-2	-2	-1	-1	-1	-0	1	-5.4	
Skin Cell (Diameter)	-8	-8	-7	-7	-7	-6	-6	**-5**	**-5**	**-5**	-4	-4	-3	-3	-3	-2	-2	-1	1	-4.3	
Nucleus of a Cell (Diameter)	-11	-9	-9	-9	-9	-8	-7	-7	-7	**-6**	-5	-5	-4	-3	-2	-1	-1	-1	1	-5.5	
Influenza Virus (Diameter)	-15	-14	-11	-10	-9	**-7**	-6	-6	-6	-6	-5	-5	-5	-5	-5	-4	-3	-1	3	-6.0	
Ribosome (Length)	-11	**-8**	**-8**	-7	-6	-5	-5	-5	-4	-3	-2	-2	-1	-1	-1	-1	-0	3	4	-3.5	
Glucose Molecule (Length)	-10	-10	-10	**-9**	-8	-8	-8	-7	-7	-6	-5	-5	-5	-5	-5	-3	-0	1	4	-5.1	
Carbon Dioxide Molecule (Length)	-9	-9	-9	-9	-8	-7	-7	-6	-6	-5	-5	-5	-4	-4	-3	-2	-1	-1	0	-5.1	

F.6 Posttest relative size values for VR participating students from the 5th through 95th percentiles.

Entity	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	M	STD
Sun (Diameter)	5	7	8	8	8	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	8.1	2.2
Saturn (Diameter)	-3	1	2	4	6	6	6	**8**	9	10	10	10	11	12	13	14	14	14	16	7.9	7.3
Venus (Diameter)	0	3	3	4	6	**7**	**7**	**7**	**7**	**7**	8	8	8	8	8	8	8	8	8	5.9	3.5
Mercury (Diameter)	2	2	3	3	4	5	5	5	5	5	**6**	**6**	**6**	**6**	**6**	8	10	13	16	6.5	4.8
Panama Canal (Length)	-1	-1	-1	-1	0	1	2	2	4	6	6	6	6	6	7	7	7	8	8	3.7	3.8
Mount Everest (Height)	-1	1	2	**4**	5	5	5	6	6	6	7	7	8	8	8	8	8	8	9	5.6	3.5
Golden Gate Bridge (Length)	-1	1	2	2	2	**3**	**3**	**3**	**3**	**3**	**3**	**3**	4	4	4	5	5	7	7	3.2	2.7
Titanic (Length)	**2**	**2**	**2**	**2**	**2**	**2**	**2**	3	3	3	3	3	4	4	5	5	5	6	6	3.5	1.8
School Bus (Length)	-5	-4	-4	-3	-1	0	0	0	0	0	**1**	**1**	**1**	**1**	**1**	2	2	3	5	0.0	3.3
Human (Height)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Squirrel (Height)	-7	-6	-5	-5	-4	-3	-3	-2	-2	-2	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	-0	1	-2.4	2.8
Bumblebee (Length)	-6	-5	-4	-4	-3	-3	-3	-3	-3	-3	**-2**	**-2**	**-2**	**-2**	-1	-1	-1	-1	3	-2.1	3.5
Fruit Fly (Length)	-5	-5	-5	-5	-5	-4	**-3**	**-3**	**-3**	-2	-2	-2	-2	-1	0	1	2	3	3	-1.9	3.1
Human Egg Cell (Diameter)	-11	-10	-9	-9	-8	-7	-7	-7	-6	-5	-5	-5	-5	-5	-5	**-4**	-3	-1	-0	-5.8	3.6
Skin Cell (Diameter)	-6	-6	-6	-6	**-5**	**-5**	**-5**	**-5**	**-5**	-4	-3	-3	-3	-3	-2	-1	0	1	1	-3.4	2.7
Nucleus of a Cell (Diameter)	-11	-10	-9	-9	-8	-8	-8	-7	-7	-7	**-6**	**-6**	-5	-4	-3	-2	-2	-1	2	-5.5	5.0
Influenza Virus (Diameter)	-10	-9	-9	-9	-9	-8	-8	**-7**	**-7**	-6	-4	-3	-2	-2	-1	0	2	3	6	-4.0	6.1
Ribosome (Length)	-11	-10	-10	-10	-10	-10	-10	**-8**	**-8**	-7	-6	-6	-5	-5	-5	-5	-5	-5	-4	-7.3	2.9
Glucose Molecule (Length)	-13	-12	-11	-11	-11	-10	**-9**	**-9**	**-9**	-8	-8	-8	-6	-6	-5	-4	-4	-3	-3	-7.9	3.8
Carbon Dioxide Molecule (Length)	-13	-12	-11	**-10**	-9	-8	-8	-8	-8	-7	-6	-5	-5	-5	-5	-4	-4	-3	2	-6.3	5.6

F.7 Pretest relative size values for non-VR participating students from the 5th through 95th percentiles.

Entity	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	M	STD
Sun (Diameter)	6	7	7	8	8	8	8	8	8	8	8	8	**9**	**9**	**9**	**9**	10	10	11	8.4	
Saturn (Diameter)	3	3	3	3	3	4	5	6	7	**8**	**8**	**8**	**8**	**8**	**8**	10	12	13	15	7.8	
Venus (Diameter)	2	2	2	2	2	2	2	3	3	3	3	3	4	4	4	5	5	6	6	3.6	
Mercury (Diameter)	4	4	5	5	5	**6**	7	7	8	9	10	11	12	13	14	14	15	15	16	9.6	
Panama Canal (Length)	-3	-2	-0	1	2	2	2	2	2	2	2	3	3	4	4	4	4	**5**	**5**	1.8	
Mount Everest (Height)	1	1	1	1	1	1	2	2	3	3	3	3	**4**	**4**	**4**	**4**	**4**	5	5	2.8	
Golden Gate Bridge (Length)	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	5	8	11	14	4.6	
Titanic (Length)	-2	-1	-1	0	1	1	**2**	**2**	3	3	3	3	4	4	4	4	4	5	5	2.0	
School Bus (Length)	-3	-2	-2	-1	0	0	0	0	0	0	0	0	**1**	**1**	**1**	2	3	3	4	0.4	
Human (Height)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Squirrel (Height)	-9	-8	-8	-7	-6	-5	-4	-4	-3	-2	-2	-2	**-1**	**-1**	**-1**	1	3	5	7	-2.0	
Bumblebee (Length)	-7	-6	-5	-4	-3	-3	-3	-3	-3	-3	-3	-3	**-2**	**-2**	**-2**	**-2**	**-2**	-1	-1	-3.4	
Fruit Fly (Length)	-8	-7	-7	-6	-6	-5	-5	-4	-4	**-3**	**-3**	-2	-2	-1	-1	-1	-1	-1	-1	-3.8	
Human Egg Cell (Diameter)	-6	-6	-5	-5	-5	-5	-5	-5	-5	-5	-5	**-4**	**-4**	-3	-3	-3	-3	-2	-2	-4.2	
Skin Cell (Diameter)	-9	-8	-8	-7	-6	-6	-6	**-5**	**-5**	**-5**	-4	-3	-1	-0	1	2	3	3	4	-3.0	
Nucleus of a Cell (Diameter)	-10	-10	-9	-9	-9	-8	-7	-7	**-6**	-5	-3	-1	0	2	4	4	4	5	5	-3.0	
Influenza Virus (Diameter)	-11	-10	-10	-9	-9	-8	**-7**	**-7**	-6	-5	-4	-3	-3	-2	-1	-1	-1	-0	-0	-5.2	
Ribosome (Length)	-9	-9	**-8**	**-8**	-7	-7	-7	-6	-6	-6	-6	-5	-5	-4	-4	-3	-3	-2	-2	-5.6	
Glucose Molecule (Length)	-7	-7	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-5	-5	-5	-2	0	3	5	-3.2	
Carbon Dioxide Molecule (Length)	-9	-9	-8	-8	-7	-7	-6	-6	-5	-5	-5	-5	-4	-4	-4	-4	-4	-4	-4	-6.0	

F.8 Posttest relative size values for non-VR participating students from the 5th through 95th percentiles.

Entity	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	M	STD
Sun (Diameter)	7	8	8	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	**9**	11	12	14	15	10.2	3.9
Saturn (Diameter)	-1	1	2	4	5	5	6	6	7	7	7	7	**8**	**8**	**8**	**8**	9	9	10	5.6	4.6
Venus (Diameter)	4	4	4	4	4	4	4	5	5	5	5	5	6	6	6	6	6	**7**	**7**	5.2	1.3
Mercury (Diameter)	2	3	3	4	4	4	4	4	4	4	5	**6**	**6**	7	8	9	11	12	14	6.6	5.2
Panama Canal (Length)	2	2	3	3	3	3	3	3	3	3	4	6	7	9	10	10	10	11	11	5.8	4.3
Mount Everest (Height)	5	5	6	6	6	6	6	7	7	7	8	8	9	9	10	10	10	11	11	7.8	2.6
Golden Gate Bridge (Length)	1	1	2	2	2	2	2	**3**	**3**	**3**	4	4	5	5	6	6	7	7	8	4.0	2.9
Titanic (Length)	2	2	3	3	3	3	3	3	3	3	3	4	4	5	5	8	10	13	15	6.2	6.7
School Bus (Length)	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	4	7	9	12	3.8	6.3
Human (Height)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Squirrel (Height)	-3	-3	-2	-2	-2	-2	-2	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	**-1**	-1.6	0.9
Bumblebee (Length)	-3	-3	-3	-3	-3	-3	-3	**-2**	**-2**	**-2**	**-2**	**-2**	-1	-1	-1	1	2	4	5	-0.4	4.2
Fruit Fly (Length)	-7	-6	-6	-5	-4	-4	-4	**-3**	**-3**	**-3**	**-3**	**-3**	-2	-2	-2	-2	-1	-1	-0	-3.4	3.0
Human Egg Cell (Diameter)	-5	-5	-5	-5	-5	-5	-5	**-4**	**-4**	**-4**	**-4**	**-4**	-3	-3	-3	-3	-3	-2	-2	-3.8	1.3
Skin Cell (Diameter)	-7	-7	-6	-6	-6	-6	-6	-6	-6	-6	**-5**	**-5**	-4	-4	-3	-3	-2	-2	-1	-4.6	2.5
Nucleus of a Cell (Diameter)	-8	-8	-7	-7	-7	-7	**-6**	**-6**	-5	-5	-5	-5	-5	-5	-5	-4	-3	-2	-1	-5.0	3.1
Influenza Virus (Diameter)	-10	-9	-8	**-7**	-6	-6	-6	-6	-6	-6	-5	-5	-4	-4	-3	-1	0	2	3	-4.2	5.9
Ribosome (Length)	-10	-10	-10	-10	-10	-9	**-8**	**-8**	-7	-6	-6	-5	-5	-4	-4	-2	-0	1	3	-5.0	6.2
Glucose Molecule (Length)	-8	-8	-7	-7	-7	-7	-7	-7	-7	-7	-6	-4	-3	-1	0	0	1	1	2	-4.0	4.6
Carbon Dioxide Molecule (Length)	-12	-11	-11	**-10**	**-10**	-9	-8	-8	-7	-6	-6	-6	-6	-6	-6	-6	-5	-5	-4	-7.6	3.3