

ABSTRACT

CHESNUTT, KATHERINE MCGREGOR. Middle Grades Students' Concepts of Size and Scale: Relationships and Factors of Influence. (Under the direction of Melissa Gail Jones.)

National K-12 science standards have included an emphasis on broad themes in science as a way to unify seemingly disparate science content (American Association for the Advancement of Science [AAAS], 1994; National Research Council [NRC], 1996; 2012). As students construct meaning by making connections between prior knowledge and new experiences, these broad themes provide a framework on which to base these connections, allowing for deep learning to occur. One such unifying theme is scale, proportion and quantity (NRC, 2012; Next Generation Science Standards [NGSS] Lead States, 2013). The purpose of this study was to examine factors related to students' concepts of size and scale. The first study includes a review of the literature on individuals' innate sense of quantity or number and the role that this innate sense plays in academic achievement. The second study reports on the relationship between students' (N=229) understanding of size and scale and students' achievement in science and mathematics. Findings indicated that a positive and significant relationship existed between students' concepts of scale and students' mathematics and science achievement in grades five through eight. The third study examines the degree to which factors such as students' out-of-school experiences with size and scale, innate sense of quantity, exposure to size and scale instruction, gender identity, and racial/ethnic identity predict students' (N=232) concepts of scale. Findings indicated that students' experiences with scale outside of school, students' racial/ethnic identities, and students' exposure to size and scale instruction statistically and significantly added to the prediction model. Implications for future research and pedagogy are discussed.

© Copyright 2017 Katherine Chesnutt

All Rights Reserved

Middle Grades Students' Concepts of Size and Scale: Relationships and Factors of Influence

by
Katherine McGregor Chesnutt

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

Science Education

Raleigh, North Carolina
2017

APPROVED BY:

M. Gail Jones
Committee Chair

Eric N. Wiebe

Sarah Carrier

Sarah Spayd

DEDICATION

To Zack.

BIOGRAPHY

Katherine McGregor Chesnutt was born in Blowing Rock, North Carolina in March of 1986, which remained her home until leaving to begin undergraduate studies at NC State. Her childhood was filled with playing in the outdoors in a small tight-knit community. The teaching profession was not something she had seriously considered, but following in the footsteps of her mentor and soccer coach, decided it could be a good fit. Following undergraduate as a Teaching Fellow at NCSU, she returned to her own high school alma mater to teach for nearly 6 years. During that time, she continued her studies and completed her MS in science education at Montana State University in Bozeman, MT, where she spent her summers. Following marriage and a move back to the triangle, she started as a full-time PhD student in STEM Education at NCSU. During that time she was met with not only a remarkable cohort of students but also extraordinary mentors. She hopes to stay closely connected to NCSU in the future as she takes time to focus on farm and family.

ACKNOWLEDGMENTS

More than anything, this process has illustrated the necessity of group cooperation and teamwork. I am eternally grateful to the many people who were essential in my success.

Dr. Gail Jones, my advisor, chairperson, and hero, without whom I would not have been able to complete this dissertation. Her extraordinary work ethic, patience, and understanding set a very high bar. She is so giving of herself that it is clear her passion about science education runs much deeper than her own degrees and professional accolades – her life’s purpose is providing meaningful education to students of all ages, and for that I greatly admire her.

My committee members, Dr. Eric Wiebe for always making time for conversations (relevant and otherwise), Dr. Sarah Carrier for her unwavering optimism, and Dr. Spayd for being willing to try something new and join the education team, even as you sail off into retirement.

My fellow graduate students, Megan, Elysa, Alonzo, Osman, Pam, Rebecca, Katie, Gina, Kylie, and Jim who are some of the few individuals that understand the ups and downs of graduate student life and were always happy to provide support. Emily for your snarky edits and making my commute into Raleigh more bearable.

The teachers, administrators and volunteers without whom this project never would have gotten off the ground and the students who were willing to show us their meter for three years in a row.

My family for being ever-present in my life. Keith for always having a couch when I needed to be in Raleigh on early mornings, and Adam for accommodating one more

Chesnutt. LP for understanding ambitions of both career and family. My parents for all of your support my entire life. Jerie for making sure I made finishing a priority and Alex for saving me a parking place. Zack for encouraging me in whatever I do. Trudy and Lily for providing canine therapy. Without all of your support, none of this would be possible.

TABLE OF CONTENTS

LIST OF TABLES.....	1
LIST OF FIGURES	2
INTRODUCTION	2
REFERENCES	4
THE APPROXIMATE NUMBER SYSTEM, A BASIS FOR THINKING ABOUT QUANTITY: EVIDENCE, MEASUREMENT, AND IMPLICATIONS FOR EDUCATION	1
ABSTRACT	1
THEORETICAL BACKGROUND	3
METHOD	5
RESULTS	7
RELATIONSHIPS BETWEEN ANS ACUITY AND ACHIEVEMENT IN MATHEMATICS.....	21
<i>Evidence of a significant relationship.....</i>	<i>21</i>
<i>Studies with absence of evidence of significant relationship between the ANS and achievement.....</i>	<i>26</i>
DISCUSSION	34
REFERENCES	37
CROSSCUTTING CONCEPTS AND ACHIEVEMENT: IS A SENSE OF SIZE AND SCALE RELATED TO ACHIEVEMENT IN SCIENCE AND MATHEMATICS?.....	1
ABSTRACT	1

LITERATURE REVIEW	4
METHODOLOGY	16
<i>Analyses.</i>	20
RESULTS	23
DISCUSSION	25
REFERENCES	37
APPENDICES	46
<i>Appendix A: Scale Card Sort</i>	47
<i>Appendix B: Comparison of Student Mean Scores Between Students that Completed Less than Half of All Assessments and Students that Completed at Least Half of All Assessments</i>	52
<i>Appendix C: Descriptive statistics for all assessments in all years of the study for all students</i>	57
STUDENTS' CONCEPTS OF SIZE AND SCALE: FACTORS OF INFLUENCE	1
ABSTRACT	1
LITERATURE REVIEW	3
METHODOLOGY	22
<i>Instruments.</i>	28
RESULTS	35
DISCUSSION	38
<i>Non-significant Factors.</i>	39
<i>Significant Factors.</i>	41

FUTURE RESEARCH AND IMPLICATIONS FOR EDUCATION	44
REFERENCES	46
APPENDICES.....	57
<i>Appendix A: Experiences with Size and Scale Survey (ESSS)</i>	58
<i>Appendix B: Tool Use Access by Percentage on the ESSS</i>	60
CONCLUSION.....	1

LIST OF TABLES

CROSSCUTTING CONCEPTS AND ACHIEVEMENT: IS A SENSE OF SIZE AND SCALE RELATED TO ACHIEVEMENT IN SCIENCE AND MATHEMATICS?

Table 1 <i>Participant sample size for each assessment across all grades</i>	18
Table 2 <i>Administration of each instrument during the multi-year study</i>	20
Table 3 <i>Pearson's correlation coefficients for the relationship between achievement and concepts of scale. Achievement is organized by content area (science or mathematics) and grade level and concepts of scale are organized by assessment constructs (relative or absolute)</i>	25
Table 4 <i>Descriptions of the North Carolina Essential Standards for science in grades 5 and 8</i>	30

STUDENTS' CONCEPTS OF SIZE AND SCALE: FACTORS OF INFLUENCE

Table 1 <i>Participant frequencies by gender and race/ethnic group</i>	23
Table 2 <i>Demographic information for control and experiment groups</i>	24
Table 3 <i>Descriptions of activities done with the experiment group across all years of the study</i>	26
Table 4 <i>Regression coefficients and standard errors for variables in the pooled data. Standardized coefficients (β) were not available using SPSS software</i>	36
Table 5 <i>Regression coefficients and standard errors for variables in each of the five Imputations</i>	37

LIST OF FIGURES

THE APPROXIMATE NUMBER SYSTEM, A BASIS FOR THINKING ABOUT
QUANTITY: EVIDENCE, MEASUREMENT, AND IMPLICATIONS FOR EDUCATION

Figure 1. An example of a non-symbolic numerical comparison task modified from
the Panamath instrument.17

STUDENTS' CONCEPTS OF SIZE AND SCALE: FACTORS OF INFLUENCE

Figure 1. The model proposed for how individuals might gain access to science scale
capital.....14

Figure 2. An example from the *Panamath* instrument.....29

Figure 3. Summary of missing values prior to imputation.....32

Figure 4. Patterns of missing values produced in SPSS.33

CONCLUSION

Figure 1. A model for factors that influence students' concepts of scale2

Introduction

This dissertation is organized into three articles, each of which describes the overall investigation of the dissertation study. The first article examined the Approximate Number System (ANS), a system in the brain that supports the rapid, imprecise ability to distinguish between large quantities (Dehaene, 2009; Halberda, Mazocco, & Feigenson, 2008). The ANS is of interest to educators because precision in individuals' ANS has demonstrated positive, significant relationships with learned skills such as formal mathematics. Though the ANS has been widely researched by neuropsychologists (Feigenson, Libertus, & Halberda, 2013), educators have yet to incorporate these findings into new pedagogies for science and mathematics. The review of the literature uses an educational perspective to examine the relationship between the ANS and academic achievement.

The second article studied the relationship between students' understanding of size and scale and students' science and mathematics achievement. Research exists on students' conceptual understanding of size and scale within the context of science (for a review see Jones & Taylor, 2009), but less is known about the relationship between students' understanding of size and scale and students' academic achievement. The Next Generation Science Standards (NGSS), which suggest teaching using broad themes such as size and scale, continues to move to the forefront of science education policy and practice (NGSS Lead States, 2013). Coupled with this movement, is the continued reliance by public schools on student academic achievement as an index of student, teacher, school, and district success. The findings of this study provided empirical evidence of the relationship between students'

concepts of size and scale and students' academic achievement. Implications for education are included.

The third article investigated the degree to which factors such as out of school experiences, innate sense of number, exposure to size and scale instruction, gender identity, and racial/ethnic identity explain individual variation in students' concepts of size and scale. Though prior research has examined the aforementioned factors as they related to size and scale independent of one another, an examination of these factors collectively has not been done. A discussion of the factors that may influence the teaching and learning of size and scale in science are included. In the chapters that follow, the proposed separate articles are identified by title, followed by a brief abstract.

References

- Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences*, 1156(1), 232-259. doi:10.1111/j.1749-6632.2009.04469.x
- Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, 7(2), 74-79. doi:10.1111/cdep.12019
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 455(7213), 665-668. doi:10.1038/nature07246
- Jones, M. G., & Taylor, A. R. (2009). Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching*, 46(4), 460-475. doi:10.1002/tea.20288

NGSS Lead States. (2013). *Next generation science standards : For states, by states*.

Washington, US: National Academies Press. Retrieved from

<http://site.ebrary.com/lib/ncsu/docDetail.action?docID=10863742&ppg=1>

The Approximate Number System, A Basis for Thinking About Quantity: Evidence, Measurement, and Implications for Education

Abstract

Recent research has shown that individuals are born with mathematical abilities that allow them to very quickly process and compare quantities of objects in an environment. This innate ability stems from a system in the brain called the Approximate Number System (ANS) that is specifically designated for processing large, imprecise quantities across multiple nonverbal modalities (Dehaene, 2009; Halberda, Mazocco, & Feigenson, 2008). Scholars largely agree that observations of the ANS in both infants and non-human animals present evidence of an evolutionarily-based ability that is a precursor to formal mathematical reasoning, researchers and educators have yet to apply knowledge of the ANS to new models of teaching and learning science and mathematics. This review of the literature describes the research on the potential relationship between the ANS and achievement in mathematics and presents the findings within an educational framework.

The Approximate Number System, A Basis for Thinking About Quantity: Evidence, Measurement, and Implications for Education

Previous studies have shown that individuals are born with mathematical abilities that allow them to rapidly discriminate between quantities. This innate ability to process numbers is supported by a system in the brain called the Approximate Number System (ANS). The ANS refers to a system in the brain specifically designated for processing large, imprecise quantities across multiple nonverbal modalities (Dehaene, 2009; Halberda, Mazocco, & Feigenson, 2008). The ANS is located in the horizontal segment of the intraparietal sulcus (HIPS), in humans and monkeys (Cappelletti, Muggleton & Walsh, 2009; Castelli, Glaser & Butterworth, 2006; Nieder & Dehaene, 2009; Piazza & Izard, 2009). Due to the presence of the ANS in both pre-verbal infants (Gallistel & Gelman, 1992) and non-human animals (Geary, 1995), there is consensus among researchers that the ANS is evolutionarily based. Additionally, recent research suggests that the ANS is a precursor to and potentially a foundation for more formal mathematical reasoning (Libertus, Feigenson, & Halberda, 2011, 2013; Libertus, Odic, & Halberda, 2012). It is likely that the ANS is related to “number sense,” a term that is not well defined but which nonetheless appears in various US mathematics standards documents (e.g., National Council of Teachers of Mathematics, 2000; National Governors Association Center for Best Practices, Council of Chief State School Officers [NGACBP, CCSSO], 2010). In the sections that follow, research on the ANS with implications for teaching and learning are reviewed, analyzed and synthesized. The overarching goal of this paper is to bring this body of research to educational researchers to gain a better understanding of what is known about the ANS in the context of

education, promote discourse, and encourage new applications of the ANS for learning mathematics and science.

Theoretical Background

From early in human history we have had to make use of mathematics for everyday survival and functioning. Although primitive mathematics skills emerged early in human history, formal mathematics is a relatively recent development (Boyer & Merzbach, 2011). Only humans in advanced societies have used complex number systems and formal mathematics to make sense of the world around them (Spelke, 2005). Yet, evidence of mathematical thinking has been observed in infants and adults with no formal instruction in mathematics as well as non-human animals (Feigenson, Libertus, & Halberda, 2013). This body of evidence has led researchers to conclude that there are parts of the brain that innately deal with diverse types of mathematical reasoning.

The two most extensively researched innate systems for mathematical reasoning are the Exact Number System (ENS) and the Approximate Number System (ANS). The ENS (also called the precise number system) is manifested in infants as an object tracking system (Piazza & Izard, 2009) or object file system (Feigenson & Carey, 2003), which allows for the monitoring of small, specific quantities of objects (Feigenson, Dehaene, & Spelke, 2004). Two lines of research showed that infants could only successfully identify up to three objects. Researchers hid different numbers of graham crackers in two opaque containers while 12-month old infants watched, and then observed which container the infants selected (Feigenson, Carey, & Hauser, 2002). Infants selected the container with the greater amount significantly more for quantities of 1 versus 2 and 2 versus 3. However, infants randomly selected containers for larger quantities

with similar ratios (e.g., 2 versus 4). Manual search experiments, in which infants searched for items that were hidden in a box, also highlighted the three-item limit (Feigenson & Carey, 2003). When infants observed 1, 2, or 3, balls being hidden in a box, the infants continued the search until the correct number of balls had been retrieved, but when 4 balls were hidden in the box, infants failed to retrieve all four balls, with most infants ending their search after retrieving only 2 of the 4 balls (Feigenson & Carey, 2003). As children learn to count, the object file systems or object tracking systems observed in infants appear to map onto the language and symbols of formal mathematics, suggesting that the ENS is essential to mathematics achievement (Libertus, 2015; Castronovo & Göbel, 2012).

The ANS, described more thoroughly below, refers to a system in the brain specifically designated for processing large, imprecise quantities across multiple nonverbal modalities (Dehaene, 2009; Halberda, Mazocco, & Feigenson, 2008). Early research on the ANS described the presence and characteristics of the ANS in infants, adults, and non-human animal species. More recently, research has focused on the relationship between individual differences in ANS acuity and students' mathematics achievement, but despite the close ties to education, these experiments have not been widely published in education journals. While the relationship between innate and learned abilities is not new to education researchers (e.g., Geary, 1995; Paas & Sweller, 2012), this review may help to promote educational research on leveraging the ANS for teaching mathematics and science.

The ANS and ENS are distinct, yet closely connected to mathematics achievement. Libertus (2015) argued that a reciprocal relationship exists between the two systems. For example, strengthening the ENS has improved ANS acuity (Castronovo & Göbel, 2012), and

individual differences in ANS acuity are correlated with future mathematics achievement (Halberda et al., 2008). However, the specific mechanisms of how the ANS and ENS interact remain unclear and are only beginning to be studied. It seems self-evident that a strong ENS might underlie a more acute understanding of symbolic number systems, leading to high mathematics achievement. In contrast, the relationship between non-symbolic, approximate number and mathematics achievement is less obvious and the significance to education remains unclear. While some scholars have argued that the ANS provides the foundation on which formal mathematical knowledge is built (Halberda & Feigenson, 2008), others have argued that skills affiliated with the ANS and formal mathematics develop independently of each other (Holloway & Ansari, 2009). This relationship of ANS to formal mathematics skills has come under increasing debate (Gilmore et al., 2013; Mussolin, Nys, Leybaert, & Content, 2016). For this reason, the possible relationship between the ANS and mathematical achievement is in need of review. The authors' interest in the ANS emerged out of research in individual differences in concepts of size (i.e., magnitude) and scale (i.e., a system for comparing magnitude), concepts that are often indexed through estimation activities. Understanding the ANS may help inform future research in studying how individuals conceptualize and approximate the size and scale of scientific phenomena.

Method

This review of the literature examined, evaluated, and integrated the existing research on innate systems of number, specifically the ANS and the role that the ANS plays in students' mathematics achievement at any age. Included in the review are both empirical and theoretical papers.

Selection of the Studies. The search for relevant research had two components: papers published on the ANS in education research journals, and papers published in cognitive psychology, neuroscience, and related fields that referenced education. Publications in education research journals on ANS were identified using Google Scholar with the keywords: “achievement,” “mathematics,” “education,” and “approximate number system,” and with journal titles including the word “education.” This search yielded 48 articles, the most relevant of which were published in two key journals: *Trends in Neuroscience and Education* (12 articles) and *Mind, Brain, and Education* (4 articles).

The search for papers in cognitive psychology and neuroscience employed a two-phase snowball technique. The earliest, highly-cited paper on ANS was identified, again using Google Scholar: an article published in the journal *Nature* by Halberda, Mazocco, and Feigenson (2008) titled “Individual Differences in Nonverbal Number Acuity Correlate with Maths Achievement.” A search was then conducted among the articles cited by this seminal paper, and the more recent publications that cited the seminal paper, using the following search terms: achievement, mathematics, education, comparison OR comparative, and “discrimination task.” The terms comparison/comparative and discrimination task were included because Halberda and colleagues (2008) used a discrimination task comparing quantities of objects to measure ANS acuity. Articles were eliminated that failed to address innate systems of number or number sense, defined by Dehaene, Dehaene-Lambertz, and Cohen (1998) as the knowledge about numbers and the relationships between numbers. This search produced 39 references that were available for analysis.

The second phase of the snowball search involved searching for pertinent references in the bibliographies of the 39 papers identified in the first phase, and the papers that cited these 39 papers. Using Google Scholar, the cited articles were examined using the same search parameters as for the first phase of the snowball search. This added an additional 42 publications. In total, the snowball search yielded 81 publications that are included in the review.

The 48 articles on the ANS in education research journals, and 81 articles on the ANS and education in journals in other fields constitute a representative and focused sample of the research on the ANS that is relevant to educational research. The review of the literature begins with a summary of research regarding the evolutionary nature of how the brain processes number, followed by an overview of the core systems of number, both precise and approximate. The final section provides a review of publications that examined the relationship between the ANS and mathematics achievement. This section on mathematics achievement is outlined thematically based on the findings reported in the literature as well as the demographics (e.g., age, dyscalculia) of the sample population. The final section also includes a summary of reviews that reflect on the efficacy of integrating neuroscience research and education.

Analysis of the studies. All 129 papers were read and analytical summaries of each were prepared. The following information was abstracted from each paper: authors, date, type of paper (theoretical or empirical), methodology, definitions of key concepts, and key findings. These summaries were then examined to detect broad overarching themes.

Results

Information processing in the brain. Scientists have grappled for years with how the

brain processes information to make meaning of the natural world. One common approach used to study the processing of information in the brain compared performance in people with healthy brains to those with brains that have cognitive-impairing lesions. A pioneering example is the case of Broca's studies of a patient who understood language but could not speak. Broca's autopsy of the patient's brain revealed lesions in the frontal lobe of the left hemisphere, in an area now known as Broca's area. Damasio and colleagues (H. Damasio, Grabowski, Tranel, Hichwa, & A. Damasio, 1996) compared word-retrieval processes in patients with normal brains to those with brain lesions and found that different regions of the brain processed specific types of words. Hills, Rapp and Caramazza (1999) found evidence to support these findings in a case study of a patient with brain lesions who demonstrated differences in semantic errors between nouns and verbs for written and spoken tasks. The patient could identify an object correctly when asked to say it, but when asked to write the object, the patient repeatedly wrote "tulip" instead of "rose" despite having said "rose" only moments before (Hillis et al., 1999).

A similar process has been used to study patients with acalculia, an acquired condition due to stroke or other brain damage that impairs mathematical ability (Ardila & Rosselli, 2002).

Macaruso, McCloskey, and Aliminosa (1993) studied the ability of a patient with brain damage who exhibited impaired numerical processing to navigate between forms of number representation (e.g., Arabic to verbal and vice versa). They found that numerical processing for specific tasks took place in distinct regions of the brain yet relied on common systematic representations of numbers. Lemer and colleagues (Lemer, Dehaene, Spelke, & Cohen, 2003) showed that different forms of acalculia exist, including quantity deficit acalculia and verbal deficit acalculia. The patient with quantity deficit acalculia could accurately multiply but

struggled to subtract, approximate quantities, and compare amounts on both symbolic and non-symbolic tasks. The patient with verbal deficit acalculia, on the other hand, struggled to subtract, but could accurately multiply, approximate quantities, and compare amounts on non-symbolic tasks (Lemer et al., 2003). These and other similar studies led to the widely accepted idea that “knowledge of different categories of words and objects such as persons, tools, animals and actions can be dissociated in brain-lesioned patients and is associated with distinct patterns of brain activation” in healthy and lesioned persons (Dehaene et al., 1998, p. 355).

In addition to behavioral tasks, brain imaging has also highlighted brain systems associated with number (Dehaene et al., 1998). Brain imaging techniques are distinct from previously described tasks in that they do not require participants to demonstrate explicit behavior (e.g., crawling, pointing, looking) and are especially helpful when working with young children (Shusterman, Slusser, Halberda, & Odic, 2016). Two main approaches to measuring brain activity are functional magnetic resonance imaging (fMRI) and event-related potentials (ERPs). According to Libertus, Woldorff, and Brannon (2007), fMRI can serve as an effective means of locating the region of the brain for specific activities, while ERPs provide a more detailed understanding of the timing of these activities. Number representations in the brain are not limited to humans, as representations of numerosity have been observed in the neuroimaging of animal brains such as cats (Thompson, Mayers, Robertson, & Patterson, 1970) and macaques (Nieder, Freedman, & Miller, 2002). Researchers found that monkeys (Piazza & Izard, 2009) and both children and adults (Dehaene, 2009; Piazza & Izard, 2009) all demonstrated activity in the mid intraparietal sulcus (on the lateral surface of the parietal lobe of the brain) during activities involving approximate number (see also Libertus et al., 2007; Piazza, 2010). Lemer,

Dehaene, Spelke, and Cohen (2003) argued that animals, infants, and adults share both homologous structures at the cellular level and the ability to approximately discriminate between amounts prior to verbal language acquisition (specifically number words). This body of research provided evidence that humans and some non-human animals shared specific networks within the brain that house different domains of knowledge (Damasio et al., 1996; Dehaene et al., 1998; Gallistel & Gelman, 1992; Hillis et al., 1999; Macaruso et al., 1993). Dehaene et al. (1998) argued that “number sense” constitutes one such knowledge domain. Other knowledge domains include the understanding of spatial relationships and the physical movements of objects or people (Spelke, 1994) as well as the comprehension and production of spoken language (Pinker, 1995). All of these knowledge domains are plausibly related to functions essential to survival and/or reproduction and are presumably the result of evolutionary processes (Geary, 1995).

Mathematical thinking and number sense. Number sense is a term with some degree of ambiguity in the mathematics education literature. The idea of number sense arose from the concept of *numeracy*, coined in 1959 by the Crowther Report in the United Kingdom (Crowther, 1959). The term numeracy was initially intended to describe an individual’s ability to navigate through modern mathematical demands and was closely tied to mathematics literacy (Crowther, 1959). However, the term numeracy was soon relegated to refer to only basic mathematical skills necessary to cope with day-to-day life (McIntosh, Reys, & Reys, 1992). The analogy of numeracy as a form of mathematical literacy continued as developments in technology led to increases in data-driven decision making that changed the way people interacted with numbers (Steen, 2001). A shift in phrasing from numeracy to being “at home with numbers” or “number sense” emerged in the 1980s and 1990s and became widely used among mathematics educators

(Hope, 1989; Howden, 1989). However, these phrases remained nearly as ambiguous as numeracy. McIntosh et al. described number sense as “a person’s general understanding of number and operations along with the ability and inclination to use this understanding in flexible ways to make mathematical judgments and to develop useful strategies for handling numbers and operations” (1992, p. 3). Number sense is highly personal and allows for the communication, processing, and interpretation of information involving numbers (McIntosh et al., 1992). In a framework for understanding basic number sense, McIntosh et al. (1992) differentiated between three key areas: number concepts, operations, and applications. McIntosh et al. (1992) argued that these three areas formed interconnections linking number sense with metacognitive processing. The authors, however, questioned whether the third component of their model, applications, differed from general problem solving and called for increasing the dialogue in this area (McIntosh et al., 1992). Cognitive neuroscientists later adopted and broadened the definition of number sense as they sought to uncover the representations that lay the foundation for numerical thought and reasoning. This review adopts the definition of number sense outlined by Dehaene et al. (1998), which refers to number sense as “the understanding of quantities and their inter-relations” (p. 355).

While complex mathematical abilities are culturally derived, number sense may be “one of the best-validated candidates for a biologically determined, domain-specific ability” (Dehaene et al., 1998, p. 355). Dehaene argued further that number is a prime candidate to study innate, domain-specific abilities because number is one of the only properties that can remain unchanged within a set, as the objects within the set change (1998). For example in a set of sounds, objects, people, or events the composition of the set can change while the number of

items remains constant (i.e., three sounds, three objects, three people, or three events) (Dehaene et al., 1998). Skepticism about infants' and animals' ability to understand abstract numbers led to a robust body of experiments examining infants' and animals' number sense across a variety of representative structures (e.g., size, shape, color of objects, spatial location, audio or visual modality, and simultaneous or sequential presentation) (Barth, Mont, Lipton, & Spelke, 2005; Dehaene et al., 1998; Halberda et al., 2008). This research revealed two core systems of number that researchers believed acted as the foundation for mathematical thinking. The following sections describe each of these systems of number representation.

Evidence for two systems of number representation: ENS and ANS. Through both behavioral and neuroscience research, researchers have found evidence of two fundamental systems of number representation. Both core systems of number exist independently of cultural symbol systems like number words or Arabic numerals. Research on infants' and animals' non-symbolic number sense – where quantities are expressed with dots, sounds, or a number of objects- provided evidence of these two distinct systems (Mussolin, Mejias, & Noël, 2010). The first core system is the ENS (described above): a system for representing amounts for smaller, more precise numbers. The second is a system for representing amounts for large, approximate numbers, the ANS. For example, Xu (2003) found that infants failed to distinguish between 2 and 4 objects, but successfully distinguished between 4 and 8 objects, despite having the same ratio. The first task requires representations of precise, smaller numbers, and the second task requires distinguishing among larger numbers, without actually quantifying either the 4 or 8 objects. Infants and non-human animals appear to have a limit of three in their ENS, so they failed to distinguish between 2 and 4. On the other hand, the small quantities involved failed to

cue the ENS. Several researchers have argued that the presence of these systems in both human infants and other species of animals indicates that the two systems are evolutionarily derived rather than culturally acquired through learning (Feigenson et al., 2013; Geary, 1996).

Furthermore, brain imaging provided evidence of distinct exact and approximate number systems because the regions of the brain associated with tasks for each do not appear to overlap (for a review see Piazza, 2009). The two systems intersect to provide a foundation for mathematical ability (Libertus, 2015). The next section describes research on the approximate number system.

The Approximate Number System. Infants, adults, and non-human animals have all been the subject of research in uncovering the core system of approximate number. Researchers have tested the ability of infants to distinguish larger numbers using a variety of methods designed to insure that the infants were only responding to number and not other factors or characteristics on display. Habituation experiments are one form of experimental design used to measure ANS acuity (Feigenson et al., 2013; Xu & Spelke, 2000). This experimental design habituated infants to a specific quantity across multiple nonverbal modalities (dots, sounds, or events) after which, researchers randomly changed the quantity across several trials. For trials that did not match the habituated quantity (e.g., habituating infants to quantities of 8 and later showing a quantity of 16 and vice versa), infants gazed longer at the image or objects (Xu & Spelke, 2000). The ability to discern between two quantities depends on how different the two quantities are. For instance, it is easier to distinguish 8 objects from 16, than 8 from 9. More specifically, differentiability between two quantities appears to follow Weber's Law, which establishes that perception is logarithmic in nature (Feigenson et al., 2004). Thus, how easily two quantities can be discriminated depends on the ratio of the two quantities. The ratios that

can be distinguished change throughout development, for a given individual, becoming closer to 1 over time. The fastest improvement occurs at very young ages and only modest improvements are observed in older children and adults. For example, Xu and Spelke (2000) found that six-month-old infants could discriminate between quantities of 8 and 16 (1:2 ratio) but not 8 and 12 (2:3 ratio). Xu and Arriaga (2007) found ten-month-old infants demonstrated the ability to discriminate quantities with a 2:3 ratio but not a 4:5 ratio. This developmental trend has also been observed in older children and adults. Researchers found that the mature levels of ANS acuity, marked by the ability to discriminate between items in a ratio of 7:8 or 10:11, were reached, at the earliest, by preteen years, with most individuals reaching mature ANS acuity between ages 20 to 30 years (Halberda & Feigenson, 2008). Adults, infants, and animals all exhibit a threshold at which the discrimination between amounts is no longer reliable, with adults having ratios closer to 1 (Lemer et al., 2003).

Researchers also found individual differences in ANS acuity within sample populations. For example, Halberda et al. (2008) found that among 14-year old children, there were significant individual differences in students' ANS acuity, even after controlling for rapid lexical access and general intelligence. Halberda et al. (2012) found a wide range of individual differences among participants of the same age group, in ages ranging from 11 to 85. Individual differences in ANS acuity have also been observed in infants and young children (Bonny & Lourenco, 2013; Halberda & Feigenson, 2008; Wang, Odic, Halberda, & Feigenson, 2016; Xu & Spelke, 2000; Xu & Arriaga, 2007).

Animals such as rats, pigeons, dolphins, raccoons, parrots, monkeys, and chimpanzees have demonstrated the ability to discriminate between numbers and discern simple operations of

addition and subtraction (e.g., Meck & Church, 1983; Meck, Church, & Gibbon, 1985). The motivation for studying animal number sense is that it is an evolutionary precursor to human arithmetic abilities. Natural selection may have favored animals with a primitive ANS, who were able to better assess the number of prey animals in herd or determine which bush has more ripe berries, providing an advantage for survival.

In summary, a variety of experiments have revealed several key characteristics of the ANS. First, the ANS is multimodal (Barth et al., 2003; Feigenson et al., 2004; Xu & Spelke, 2000). Another signature of the ANS is that it represents approximations of quantities so that distinct amounts can be differentiated up to specific ratio limits that change throughout development and follow Weber's law (Halberda & Feigenson, 2008; Halberda et al., 2012). Another signature of the ANS is that precision of the ANS differs among individuals and differences exist across age groups (Halberda et al., 2008; Halberda et al., 2012). A fourth signature is that improvements in ANS acuity are not evenly spaced throughout development, with the most substantial gains in ANS precision occurring during infancy and tapering off in adulthood (Halberda & Feigenson, 2008; Halberda et al., 2012; Piazza, 2010; Xu & Spelke, 2000). Another signature is that brain imaging has demonstrated that numerical processing of the ANS is located in the intraparietal sulcus (Dehaene, 2009; Piazza & Izard, 2009). Lastly, the ANS is present in both human and non-human animal species, and in preverbal humans, strongly suggesting an evolutionary origin (Dehaene et al., 1998; Meck & Church, 1983; Meck et al., 1985).

Measuring ANS Acuity. Since ANS involves the ability to distinguish two different quantities, the key measure is how different those two quantities must be for reliable distinction

(i.e., their ratio). In addition, to avoid mobilizing the ENS rather than the ANS, the quantities presented must be larger than can be subitized. Furthermore, the time of presentation of the two quantities must be short in order to preclude counting. Methods for measuring ANS acuity are described next.

Non-symbolic comparison tasks. Procedures used to collect data on individuals' ANS acuity typically take three forms. In the first, participants must discriminate between two non-symbolic quantities such as deciding which of two collections of objects is greater in number (e.g., Halberda et al., 2008; Halberda & Feigenson, 2008). The Psychophysical Assessment of Numerical Approximation (*Panamath*) uses dots of two different colors, and the person must quickly decide what color dots are more numerous (Halberda, Ly, Wilmer, Naiman, & Germine, 2012). The online platform of *Panamath* has allowed for massive samples of over 10,000 participants. Figure 1 includes an example of a non-symbolic comparison task modified from the *Panamath* instrument. Characteristics such as area or luminosity are controlled in an effort to make numerosity the salient difference.

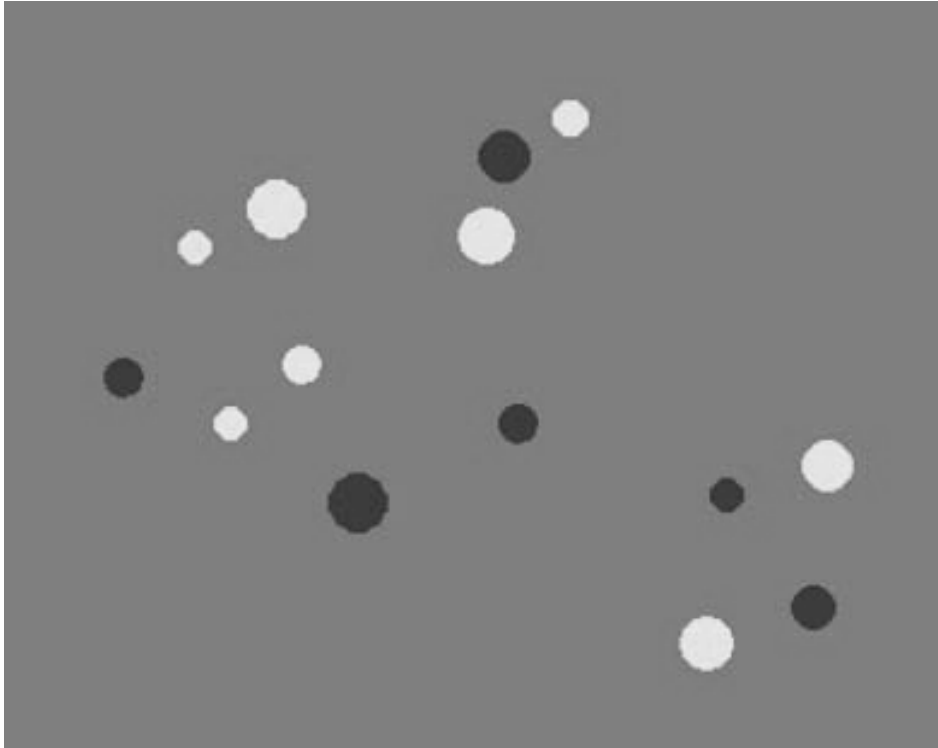


Figure 1. An example of a non-symbolic numerical comparison task modified from the *Panamath* instrument. The dots on display are normally presented in blue and yellow.

Another approach is to prompt participants to add (e.g., Barth et al., 2005; Gilmore, McCarthy, and Spelke, 2010) or subtract (e.g., Gilmore, McCarthy, and Spelke, 2007) two large sets of dots and compare the computed quantity to a third amount. To ensure activation of the ANS (rather than ENS) on addition and subtraction trials, the range of magnitudes in comparison tasks in the aforementioned experiments ranged from 5-98 (Barth et al., 2005; Gilmore et al., 2007; 2010). Gilmore et al. (2010) argued that non-symbolic addition was a more sensitive measure of future mathematical ability because it requires both comparison and transformation (i.e., addition) of amounts, which collectively lay the foundation for mathematics knowledge. Another approach to measuring ANS acuity has been to ask participants to match two non-

symbolic quantities of the same amount. For example, Nys, Ventura, Fernandes, Querido, and Leybaert (2013) asked participants to match a given quantity of dots to one of three sets of dots, ranging in quantity from 20 to 90. Other instruments and tasks have been designed to investigate the relationship between symbolic and non-symbolic quantities such as matching a quantity of dots to the corresponding two-digit Arabic numerals (Nys et al., 2013). As mentioned previously in the methodology section, for the purposes of limiting the scope of this robust area of research as well maintaining consistency across the reviewed articles, only research that examined the relationship between non-symbolic comparison tasks and mathematics achievement was considered in this review. For an evaluation of non-symbolic comparison tasks for examining ANS acuity (specifically with dots), see Clayton, Gilmore and Inglis (2015).

ANS acuity metrics. There are a number of assessments that claim to validly and reliably measure ANS acuity for non-symbolic comparison tasks. The most simple of these measures is to report the accuracy with which an individual can discriminate between quantities while controlling for time of presentation of the stimuli and ratio (e.g., Libertus, Feigenson, & Halberda, 2013; Fuhs & McNeil, 2013). Another approach to measuring ANS acuity is to examine the speed at which an individual can accurately discriminate between quantities (usually, while controlling for their ratio), referred to in the literature as reaction time (RT). Both accuracy and RT often serve as supporting data that can be used in conjunction with measures of the ratio closest to 1 that the test taker can reliably distinguish. These measures include the Weber fraction, Numerical Distance Effect (NDE), or the Numerical Ratio Effect (NRE), described next.

The Weber fraction is a value that uses the ratio, rather than difference, to characterize the two values to be discriminated (Getty, 1975; Grondin, Ouellet, & Roussel, 2001; Halberda & Feigenson, 2008). Halberda and Feigenson (2008) explained that the Weber fraction “is equal to the difference between the two numbers divided by the smaller number; for example, $7:8 \rightarrow (8-7)/7 = 0.14$,” (p. 1457). Thus, a lower Weber fraction indicates the ability to discriminate between numbers that are numerically closer, and a higher Weber fraction indicates the ability to discriminate only between numbers that are numerically farther away (e.g., $8:12 \rightarrow (12-8)/8 = 0.5$). Price, Palmer, Battista, and Ansari (2012) described the Weber fraction as “an estimation of the noise in the representation that drives that performance and can be used to predict participants’ relative magnitude comparison performance for any given ratio,” (p. 51). A large body of research has shown that ANS acuity (as measured by the Weber fraction) improves through childhood with average Weber fractions of 1.0 (corresponding to 1:2 ratio), 0.5 (2:3 ratio), and 0.14 (7:8 ratio) observed at 6 months, 9 months, adulthood, respectively (Halberda & Feigenson, 2008; Xu & Spelke, 2000; Xu & Arriaga, 2007).

Other researchers have used the Numerical Distance Effect (NDE) as a metric in comparison tasks involving numerical magnitude. The NDE uses the difference or distance between values (rather than ratio) to determine the difficulty or ease of discrimination (Dehaene, Dupoux, & Mehler, 1990; Holloway & Ansari, 2009). In the early literature, it was unclear what the methodological and analytical benefits of one comparative measure were over another. Gilmore, Attridge, and Inglis (2011) argued that because in their findings the NDE did not correlate with achievement in mathematics (e.g., Holloway & Ansari, 2009), it was not an accurate measure of ANS acuity. Further research was needed to determine if there were

significant differences in using the Weber fraction or the NDE for analyzing comparison tasks involving quantity.

In an attempt to better understand the inconsistent results of studies involving the ANS, Price, Palmer, Battista, and Ansari (2012), conducted an evaluation of both Weber fraction and the Numerical Ratio Effect (NRE). Much like the NDE, the NRE “refers to a monotonic increase in reaction time and error rate (accuracy) as the *ratio* (smaller/larger) between two numbers increases,” (Price et al., 2012, p. 51). Price et al. (2012) explained that the NRE and NDE were highly correlated and used interchangeably in the literature. Price et al. (2012) calculated individual Weber fractions and NREs across three non-symbolic tasks, historically used to measure ANS acuity. Price et al. (2012) compared their results across metrics within their study and between previous studies that used either NREs or Weber fractions. Price et al. (2012) found that the accuracy of participants’ NREs was not reliable or valid for measuring ANS acuity, but reaction time of the participants’ NREs was reliable. In contrast, the Weber fraction had stronger correlations within and between tasks than the NRE, leading the authors to suggest that the Weber fraction was a stronger measure of ANS acuity (Price et al., 2012). However, in an investigation of the indices of ANS acuity (accuracy, Weber fraction, NRE-RT, and NRE-accuracy), Inglis and Gilmore (2014) found that the distribution of participants’ Weber fractions were skewed in a positive (i.e., more precise) direction, while the distribution of accuracy, NRE-RT, and NRE-accuracy were normal (Inglis & Gilmore, 2014). Contrary to Price et al. (2012), Inglis and Gilmore (2014) argued that the best index or measure of ANS acuity was accuracy rather than Weber fraction. More recently, in a meta-analysis of research on ANS acuity, Schneider et al. (2016) found evidence in support of Price and colleagues’ (2012) claims

that the Weber fraction served as the best means of indexing ANS acuity, but this debate remains unsettled in the literature.

Recent work by Halberda and colleagues has attempted to include both reaction time and accuracy as measured by Weber fraction, by taking the average of the normalized z-scores of both measures (Authors, 2016; Feigenson, Libertus, and Halberda, 2013).

Relationships between ANS Acuity and Achievement in Mathematics

In addition to disagreement in the literature with regard to best ways of indexing ANS acuity, there have also been inconsistent results with regard to whether or not a relationship exists between ANS acuity and achievement in mathematics. The following section summarizes research that found a significant relationship between ANS acuity and mathematical achievement followed by a summary of research that found no such evidence. Each subsection (both significant and not significant) is organized based on sample populations, specifically children and adults.

Evidence of a significant relationship.

Children. In a seminal study of the relationship between the ANS and mathematics achievement, Halberda et al. (2008) examined individual differences in ANS acuity on the *Panamath* assessment, and how those differences correlated with individual differences in formal, symbolic mathematics abilities. Using the Weber fraction as a metric of comparison, Halberda et al. (2008) found that among 14-year old children, there were significant individual differences in students' ANS acuity, even after controlling for rapid lexical access (often used as a predictor of verbal ability) and general intelligence. Halberda et al. (2008) also found a significant correlation between ANS acuity and past mathematics achievement (with greater

acuity corresponding to higher achievement) over a nine-year period, suggesting that individual differences in ANS acuity likely contributed to the individual differences observed in mathematical achievement as measured by annual standardized state mathematics assessments. In a follow-up study, Halberda and Feigenson (2008) argued that because the development of the ANS occurred over several years (they found incremental improvement in ANS acuity from ages 3 to 6), the ANS may have implications for improving mathematics education as well as understanding the role that individual experiences may play in the development of the ANS. However, the authors provided no specific descriptions about the form that the proposed interventions might take, a common omission throughout the research on this relationship.

Inglis, Attridge, Batchelor, and Gilmore (2011) argued that because Halberda et al. (2008) measured 14 year olds' ANS acuity and correlated it to their past mathematics achievement, it did not take into account the environmental factors that may have influenced students' ANS acuity. Using non-symbolic dot comparison tasks, based on Pica, Lemer, Izard, and Dehaene (2004), and Weber fractions as a measure of ANS acuity, Inglis et al. (2011) concurrently measured and compared ANS acuity and mathematics ability in both children and adults. Inglis et al. (2011) found that ANS acuity correlated with mathematics achievement in children based on standardized mathematics assessments. Results for adults are discussed in the following section.

To further investigate the role of the ANS in formal mathematics, Mazzocco, Feigenson, Halberda, and Santos (2011) demonstrated that ANS acuity, measured prior to formal mathematics instruction (in preschool), predicted mathematics ability in elementary school, irrespective of vocabulary, perceptual organization, or non-numerical lexical retrieval. While

this sample size was relatively small ($N=17$), the authors argued that the findings provided evidence that individual differences in ANS acuity likely play a role in the acquisition of mathematical abilities, yet provided no insight as to how this relationship could influence instruction (Mazzocco et al, 2011).

Libertus, Feigenson, and Halberda (2013) conducted an experiment with preschool-aged children ($N=169$) to measure the relationship between ANS acuity and mathematics. The researchers improved on previous experiments in several ways. First, Libertus et al. (2013) improved on prior methods by administering the tests for both ANS acuity and mathematics skills two separate times and to a larger sample size than had been used in previous studies (e.g., Mazzocco et al., 2011). Second, the researchers tested the participants two times with a six month delay between each testing session to better understand the role of development in both ANS acuity and mathematics achievement. Lastly, the researchers controlled for initial math ability, reaction time, working memory, age, and non-numerical ability such as vocabulary. Libertus et al. (2013) found that ANS acuity measured in the first round of testing predicted growth in mathematics achievement across the six-month period, even after controlling for age, initial mathematics ability, and non-numerical ability. Similar to prior research, Libertus and colleagues (2013) claimed their research could have implications for education, yet offered no specific instructional strategies or interventions.

In another study, Bonny and Lourenco (2013) investigated the relationship in various age groups of children age 3 to 5. Previous research had investigated the relationship between ANS acuity and achievement but had not looked at the relationship within and between specific age groups. The researchers found a significant relationship between ANS acuity and mathematics

achievement for all participants, but when analyzed by age group, found a significant relationship for both 3-year-olds and 4-year-olds but not 5-year-olds (Bonny & Lourenco, 2013). However, in a study of elementary students in grades two and four, Park and Cho (2016) found a significant relationship between ANS acuity and mathematics achievement in fourth graders but not second graders. These findings led to speculation that there could be differences in the relationship between ANS acuity and mathematics throughout development and into adulthood, yet offered little insight into how these findings might influence education. For a review of the relationship between symbolic and non-symbolic number sense across development see Mussolin et al. (2016).

Adults. Similar trends in the relationship between ANS acuity and mathematical ability have also been observed in adults. In what is likely the largest study of ANS acuity to date, Halberda et al. (2012) measured more than 10,000 individuals ranging in age from 11 to 85. Among all the participants in the sample, most of them adults, Halberda et al. (2012) found a significant relationship between ANS acuity as measured by the *Panamath* instrument and participants mathematics ability (measured by self-reported quantitative scores on the Scholastic Aptitude Test [SAT]) with reaction times and Weber fractions negatively correlated with SAT scores ($r = -0.21$) (Halberda et al., 2012), indicating that higher math achievement is associated with higher ANS acuity (indicated by lower reaction times and Weber fractions).

Libertus, Odic, and Halberda (2012) also examined the relationship between ANS acuity and quantitative scores on the SAT. After controlling for SAT-verbal scores, Libertus et al. (2012) found a weak but significant correlation between mathematics scores on the SAT and ANS acuity (measured by Weber fraction and reaction time) for experiments conducted both in a

college classroom and in a controlled laboratory setting ($r = -0.19$; $p < .05$). Libertus et al. (2012) argued that the association between ANS and mathematical skills is initiated in childhood and continues to play a role into adulthood.

Nys et al. (2013) investigated the difference in ANS acuity between three groups of adults living in the same Western culture: those without formal mathematics education, those with formal education acquired as an adult, and those with formal education acquired in childhood. Results showed that the group with no formal education performed slower, with less accuracy, and had smaller NRE than the other groups on large quantity (i.e., numerosities of 20-90) comparison tasks for non-symbolic (i.e., comparing two sets of dots) and symbolic (i.e., comparing two Arabic numerals) magnitudes (Nys et al., 2013). In addition, the group with no formal education was less accurate than other groups in matching both non-symbolic and symbolic quantities (i.e., matching a set of dots to one of three Arabic numerals) as well as matching two non-symbolic amounts (i.e., matching a set of dots to one of three sets of dots) (Nys et al., 2013). Similar to Libertus et al.'s (2012) hypothesis that the ANS and the precise number system form a reciprocal relationship, Nys et al. (2013) argued that one explanation of their results could be that learning precise numbers through mathematics instruction could improve ANS acuity in adults.

Further evidence of the relationship between formal mathematics and ANS acuity was observed in children with developmental dyscalculia. Different from acalculia, which is acquired through brain injury, dyscalculia is considered an innate or developmental disability in mathematics (Ardila & Rosselli, 2002). Piazza et al. (2010) found that 10-year-old children with dyscalculia performed on level with 5-year-old children without the disability on non-symbolic

comparison tasks, demonstrating a link between dyscalculia and an impaired ANS. This was supported by findings of Mazzocco, Feigenson, & Halberda (2011) who reported that ninth grade students with dyscalculia have significantly lower ANS acuity than their non-dyscalculaic peers.

Much of the research that found significant relationships between ANS acuity and mathematics achievement emerged from research conducted by Halberda and colleagues, raising the question whether their methodology (e.g., Panamath, and Weber fraction as metric) might possibly be biased. Conversely, the large sample size of many of their studies compared to other researchers might lead to statistically significant results that other, low-power designs cannot detect. However, different approaches used to measure ANS acuity have demonstrated contrary results and are discussed in the following section.

Studies with absence of evidence of significant relationship between the ANS and achievement.

Children. In one of the earlier studies of the ANS, Holloway and Ansari (2009) used the NDE and comparison tasks to examine the relationship between students' achievement on standardized tests and both symbolic and non-symbolic NDE. Holloway and Ansari (2009) found that students with smaller distances between the symbolic numbers they could discriminate (i.e., lower NDE), had higher scores on their mathematics achievement tests. However, the same did not hold true for non-symbolic number discrimination, leading the authors to suggest that symbolic and non-symbolic NDEs leverage different systems with different relationships to mathematics achievement (Holloway & Ansari, 2009). These findings contrasted those reported by Halberda et al. (2008), yet some researchers surmised that methodological differences in measuring ANS acuity (i.e., using the Weber fraction versus using

NDE) led to the different conclusions (Price et al., 2012). It is worth noting that Weber fraction involves relative difference between numbers (i.e., a ratio), while NDE involves absolute difference (e.g., a subtractive difference). Given the widespread applicability of Weber's law in human sensory perception, it seems likely that Weber fraction is a more appropriate measure than NDE.

More recently, in a study of 1,463 Dutch children in grades 1–6, Lyons, Price, Vaessen, Blomert, and Ansari (2014) found that ANS acuity did not predict mathematics ability. Though the authors cautioned that their results did not examine the relationship between ANS acuity and mathematics ability prior to formal mathematics instruction, they argued that the claims that non-symbolic number sense should be used to guide mathematics instruction (e.g., Halberda and colleagues) are unfounded (Lyons et al., 2014). The authors found a significant relationship for all grades ($r = .554$; *effect size* = 1.331; $p < 0.001$), but when controlling for other variables such as symbolic number tasks, the authors found the relationship was weak and not significant for all grades (e.g., for sixth grade, $r = 0.021$; *effect size* = 0.043) (Lyons et al., 2014).

Fuhs and McNeil (2013) claimed that most of the leading ANS studies had been conducted with a homogenous sample population (i.e., white, upper-middle class). To address this claim, Fuhs and McNeil (2013) investigated the relationship between ANS acuity and mathematics ability in students from low-income households. When accounting for inhibitory control (the ability to suppress irrelevant information), Fuhs and McNeil (2013) found no significant relationship between ANS acuity and mathematics ability. Total ANS acuity, however, was significantly correlated with inhibitory control, particularly for area versus quantity inverse trials. In these trials, the total area of the set of dots with a lower quantity was

greater than the total area of the set of dots with a greater quantity. The mismatch between area and quantity in these trials required students to suppress the irrelevant information of area in order to focus solely on quantity. Inhibitory control becomes increasingly necessary as mathematical problem solving becomes more complex (e.g., word problems) (Fuhs & McNeil, 2013). Fuhs and McNeil (2013) argued that environmental factors before the start of formal mathematics instruction, more than the ANS, play a role in the development of mathematics ability. Fuhs and McNeil (2013) speculated that ANS acuity measures both children's ability to approximate magnitudes and their inhibitory control, with individuals that have a more precise ANS having greater inhibitory control.

Gilmore et al. (2013) reported similar findings to Fuhs and McNeil (2013). By using response times, accuracy, and the Weber fraction as a measure of ANS acuity, Gilmore et al. (2013) found that overall, ANS acuity and response times correlated with achievement in mathematics. However, when the researchers examined dot comparison trials that had inverse relationships between dot area and quantity, Gilmore et al. (2013) found that inhibitory control, rather than ANS, contributed a greater amount to the accuracy on each trial. For more on the relationship between inhibitory control and mathematics ability in students from diverse backgrounds see Ng, Tamis-Lemonda, Yashikowa, and Sze (2015).

Adults. Some research on adults' ANS acuity has also demonstrated no significant relationship between ANS precision and mathematics ability. Inglis et al. (2011) concurrently measured and compared ANS acuity and mathematics ability in both children and adults. As mentioned earlier, Inglis et al. (2011) found that ANS acuity correlated with mathematics achievement in children, but the same did not hold true for adults. Inglis et al. (2011) speculated

that because of the simplistic nature (i.e., whole numbers) of early symbolic mathematics, it is likely that the ANS played a stronger role at young ages, and as symbolic mathematics increased in abstraction into adulthood, other factors (e.g., working memory, teacher effect, strategy) weighed more heavily in determining individual differences in mathematics achievement. Findings by Inglis et al. (2011), point to the idea that if improving ANS in young children can lead to improved mathematics achievement, targeted interventions will likely be more effective at younger ages or earlier in development. However, beyond suggestions of *when* to conduct interventions, few specifics about the type of intervention were included in the discussion by Inglis et al. (2011).

In an investigation of adults (mean age 22.29 years), Price et al. (2012) used both the NRE and Weber fraction across three tests of non-symbolic number acuity (differing in dot presentation: paired simultaneous, isolated sequential, and random, non-overlapping simultaneous) and found very weak and non-significant relationships between the indices of ANS acuity and mathematics achievement as measured by the Woodcock Johnson Math Fluency subtest. This investigation included a relatively small sample size ($N=36$) compared to studies that found contrary results (e.g., Halberda et al., 2012) and made no suggestions for how this research might influence the integration of neuroscience and education.

Castronovo and Göbel (2012) found a significant correlation between symbolic (rather than non-symbolic) numerical abilities and mathematics achievement. The authors noted that these results were not contrary to findings by Halberda et al. (2008); rather, the authors suggested that their results, along with results presented by Halberda and colleagues, indicated that most individuals reach maximal ANS acuity (typically the ability to discriminate between

ratios of at least 7:8) at some point during development irrespective of mathematics instruction or achievement (Castronovo & Göbel, 2012). Similar to results found by Holloway and Ansari (2009), Castronovo and Göbel (2012) found a significant relationship between mathematics achievement and the ability to map between symbolic and non-symbolic numbers. Castronovo and Göbel (2012) argued that it is the ability to map between symbolic and non-symbolic numbers, rather than non-symbolic ANS acuity, that supports abilities in mathematics. This has led to calls by researchers for further investigations of the learning trajectory and relationship between symbolic and non-symbolic mathematics (De Smedt, Noël, Gilmore, & Ansari, 2013; Schneider et al., 2016).

Overall, the findings of the review on the relationship between ANS acuity and mathematics achievement demonstrated mixed results and possible differences due to development. Since this is likely the result of varied methodologies, metrics, and sample sizes, more research is needed in this area (Inglis & Gilmore, 2014; Price et al., 2012).

Emergent theories on ANS and mathematics achievement. The link between the ANS and mathematical abilities is far from clear. However, several leading hypotheses have emerged in the literature. One such theory emerged from research proposing that ANS serves as the foundation on which mathematical ability is built. From this view, children make meaning of mathematical symbols (i.e., Arabic numerals) when the symbols are mapped onto and form associations with the ANS (Barth et al., 2005). Later or in some cases concurrently, the mapping of symbols onto the ANS allows for a more robust understanding of number which is manifested in performance on mathematics assessments as the relationship exists throughout development and into adulthood (Halberda et al., 2008; Halberda et al., 2012). Building on this idea, the

relationship between the ANS and access to symbolic mathematics, allows children with greater ANS acuity to more easily learn to count and learn symbolic numbers, potentially allowing for a greater understanding of symbolic number representations (Libertus et al., 2012). Furthering this hypothesis is the idea that lower ANS acuity could lead to complications with performance on problems involving arithmetic and ordinal relationships, because students with low ANS acuity to rely on rigid and specific algorithms rather than critical thinking or estimation when solving and evaluating the answers to arithmetic problems (Libertus et al., 2012). Other researchers have also posited a that low ANS acuity could set off a series of relationships in which low ANS acuity leads to low engagement, which leads to high mathematics-related anxiety, which leads to a decrease in mathematics performance (Maloney, Ansari, & Fugelsang, 2011). A slightly different hypothesis is that the relationship between the ANS and mathematics achievement is reciprocal and that the confidence students gain from a foundational understanding of what is “bigger” or “smaller” has a lasting effect even as individuals learn high mathematics as adults (Libertus et al., 2012).

A second theory, contrary to the idea of mapping, has emerged from research that found no significant relationship between ANS acuity and mathematics ability. This view is based on the idea that nonsymbolic representation of magnitude does not serve as a precursor to symbolic representation of magnitude and that symbolic and non symbolic numerical processing are not related (Holloway & Ansari, 2009).

Overall, most scholars agree that a link exists between innate numerical processing and formal mathematics skills, yet the mechanism and implications of this relationship remain unclear. A recent development to investigate the ways in which ANS acuity might affect

mathematics achievement has been the introduction of intervention studies. Different from all previous studies that involved correlational analyses, controlled intervention studies open new possibilities to understanding this complex relationship.

Intervention Studies. The divergent theories outlined in the previous section led researchers to call for an increase in controlled experimental studies that could elicit causal, rather than correlational relationships. Few intervention studies met the search parameters for this review because they examined the relationship between symbolic (rather than non-symbolic) number representation and mathematics ability (for a review of symbolic intervention studies see De Smedt et al., 2013). In one of the first intervention studies on the role that education may play in ANS acuity, DeWind and Brannon (2012) conducted an experiment to measure the effects of feedback and training (i.e., informing students of the accuracy of each response) on a non-symbolic comparison task. DeWind and Brannon (2012) found participants' ANS acuity improved only after the second of six total sessions. Researchers observed rapid improvement immediately following the first session, which included feedback on each trial. However, improvement plateaued following the second session, suggesting a limit to the malleability of ANS acuity (DeWind & Brannon, 2012). In a follow-up experiment, Park and Brannon (2013) found that ANS acuity in participants could improve with intervention, and this improvement was correlated with improved mathematics ability. This experiment, however, employed a non-symbolic approximate *arithmetic* task rather than a non-symbolic *comparison* task and was, therefore, unable to demonstrate a direct causal relationship that met the parameters of this review (DeWind & Brannon, 2012).

More recently, Wang, Odic, Halberda, and Feigenson (2016) sought to examine whether changing ANS acuity could affect mathematics ability. In their own prior research, Odic and colleagues found that the order in which children completed dot comparison tasks, either progressively harder (Easy First) or progressively easier (Hard First), could affect ANS acuity (Odic, Hock, & Halberda, 2014). Despite an even distribution in ability across groups of participants, researchers observed greater ANS acuity in those students who completed the “Easy First” task sequence than those who completed the “Hard First” sequence (Odic et al., 2014). Wang et al. (2016) investigated whether or not changing the order of comparison tasks (and thus manipulating students’ ANS acuity) could affect achievement in mathematics. In one of the few causal experiments involving the ANS and mathematics, Wang et al. (2016) found that children ($N=40$, mean age of 5 years 4 months) who completed the “Easy First” ANS task, performed significantly better on the measure of mathematics achievement than the “Hard First” group, when controlling for vocabulary and verbal abilities. This study provides the best evidence of a causal relationship between ANS acuity and mathematics ability in children (Wang et al., 2016). Studies of both the correlational and causal relationships between ANS acuity and achievement in mathematics have led researchers to suggest that this area of research could have significant implications for education. However, despite these suggestions, the implications for education in all studies lack specificity that could be adopted by educators. The following section provides a summary of suggestions and challenges that arise from efforts to integrate neuroscience and education.

Discussion

Some researchers have touted the implications for how research on the ANS could change mathematics education, while others have warned that insufficient evidence exists to link the two. While there is evidence supporting both accounts, a robust amount of research increasingly points to a significant, albeit complex, relationship between the ANS and mathematical ability. Thus, there are ways in which mathematics education researchers specifically, and STEM (science, technology, and engineering educational) education researchers in general, might leverage the findings in cognitive psychology and neuroscience to improve mathematics research and instruction. To date, few successful connections have been made between the two domains. To explain this gap, some researchers pointed to the inability of educational stakeholders (e.g., parents, teachers, etc.) to understand the challenges associated with translating neuroscience research into classroom practice (Looi, Thompson, Krause, & Cohen Kadosh, 2016), while others pointed to a fundamental overreach by neuroscience researchers on the applicability of their research to classroom practice (Bowers, 2016). For example, Bowers (2016) argued that the successes claimed by researchers in the field of education-related neuroscience were trivial at best, because they were self-evident, based on well-established behavioral psychology (rather than neuroscience), and presented conclusions that misrepresented or failed to follow actual neuroscience (e.g., learning styles). Bowers noted that educational neuroscience has failed to improve instruction because “what matters is not whether the brain changes, but whether the child learns as expressed in behavior,” (2016, p. 10).

Nevertheless, not all efforts to integrate neuroscience and education have been met with criticism. Importantly, neuroscience research offers methodological approaches such as fMRI

and ERPs that can support behavioral research, as long as the research is grounded in cognitive theories (De Smedt, 2014; De Smedt & Grabner, 2015). Furthermore, neuroimaging allows for a more comprehensive and in-depth understanding of learning at the biological level, thus supporting or clarifying research in behavioral psychology and education (De Smedt, 2014; De Smedt & Grabner, 2015). De Smedt and Grabner (2015) posited that integrating neuroscience and education could provide opportunities for both “neuropredictions” and “neurointerventions” that could inform education. As an example of this in practice, De Smedt (2014) argued that neuroscience research has led the development of diagnostic instruments for detecting developmental dyscalculia that employ measures of ANS acuity.

While there has been a continued effort to integrate neuroscience and mathematics education research for the benefit of both science knowledge and educational practice, bridging this gap is not without its challenges. For example, a decisive conclusion as to the relationship between the ANS and mathematics achievement from the neuroscience community as well as further investigations on the effects of targeted ANS interventions, could help educators see merit in the findings of neuroscientists. Likewise, educators that are more informed of the current research in neuroscience may be more likely to adopt strategies derived from neuroscience research. Furthermore, there is a need to understand how brain systems that process specific domains of knowledge might be used in other content areas. For example, understanding the relationship between ANS acuity and achievement in science or engineering could help inform researchers’ understandings of the ways in which individuals make meaning out of the world with regard to science or engineering.

Paas and Sweller (2012) suggested that skills are not developed through either biologically primary (innate) or biologically secondary (culturally learned) knowledge alone; rather, skills are developed through the integration of both forms of knowledge. Thus, ANS and knowledge of number might be one such combination that may foster skills in mathematics or other content areas such as science and engineering. The relationship between innate numerical ability and students' concepts of estimation, measurement, and scale is of interest to science education researchers (Authors, 2016), but there is a need for further exploration from an educational perspective so that educators might leverage innate abilities to improve learning of culturally acquired skills. Through an interdisciplinary partnership, educational researchers and neuroscience researchers have the potential make significant advances in each of their domains but much work is needed to make it a fruitful collaboration.

References

- Ardila, A., & Rosselli, M. (2002). Acalculia and dyscalculia. *Neuropsychology Review*, *12*(4), 179-231. doi:1021343508573
- Authors. (2016). Blinded for review. Submitted for publication.
- Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201-221. doi:10.1016/S0010-0277(02)00178-6
- Barth, H., Mont, K. L., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 14116-14121. doi:10.1073/pnas.0505512102
- Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology*, *114*(3), 375-388. doi:10.1016/j.jecp.2012.09.015
- Bowers, J. S. (2016). The practical and principled problems with educational neuroscience. *Psychological Review*, doi:10.1037/rev0000025
- Boyer, C. B., & Merzbach, U. C. (2011). *A history of mathematics*. John Wiley & Sons.
- Bynner, J., & Parsons, S. (2006). Does numeracy matter more? National Research and Development Centre for Adult Literacy, Institute of Education, London.
- Cappelletti, M., Muggleton, N., & Walsh, V. (2009). Quantity without numbers and numbers without quantity in the parietal cortex. *Neuroimage*, *46*(2), 522–529.

- Castelli, F., Glaser, D. E., & Butterworth, B. (2006). Discrete and analogue quantity processing in the parietal lobe: A functional MRI study. *Proceedings of the National Academy of Sciences*, *103*(12), 4693–4698.
- Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*(1), 59-68.
doi:10.1162/001152604772746701
- Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE*, *7*(4), 1-16. doi:10.1371/journal.pone.0033832
- Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica*, *161*, 177-184.
doi:10.1016/j.actpsy.2015.09.007
- Crowther, G. (1959). A report of the central advisory council for education.
- Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature*, *380*(6574), 499-505. doi:10.1038/380499a0
- De Smedt, B. (2014). Advances in the use of neuroscience methods in research on learning and instruction. *Frontline Learning Research*, *2*(4), 7-14.
- De Smedt, B., & Grabner, R. (2015). Applications of neuroscience to mathematics education.
- De Smedt, B., Noël, M., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48-55. doi:10.1016/j.tine.2013.06.001

- Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology. Human Perception and Performance*, 16(3), 626-641.
- Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences*, 1156(1), 232-259. doi:10.1111/j.1749-6632.2009.04469.x
- Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, 21(8), 355-361. doi:10.1016/S0166-2236(98)01263-6
- DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience*, 6. doi:10.3389/fnhum.2012.00068
- Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, 6(5), 568-584. doi:10.1111/1467-7687.00313
- Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, 13(2), 150-156.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307-314. doi:10.1016/j.tics.2004.05.002
- Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, 7(2), 74-79. doi:10.1111/cdep.12019

- Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science, 16*(1), 136-148. doi:10.1111/desc.12013
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1-2), 43-74. doi:10.1016/0010-0277(92)90050-R
- Geary, D. C. (1995). Reflections of evolution and culture in children's cognition. *American Psychologist, 50*(1), 24-37.
- Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19*(02), 229-247. doi:10.1017/S0140525X00042400
- Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics, 18*(1), 1-8. doi:10.3758/BF03199358
- Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature, 447*(7144), 589-591. doi:10.1038/nature05850
- Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115*(3), 394-406. doi:10.1016/j.cognition.2010.02.002
- Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE, 8*(6) doi:10.1371/journal.pone.0067374

- Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *The Quarterly Journal of Experimental Psychology*, *64*(11), 2099-2109.
doi:10.1080/17470218.2011.574710
- Grondin, S., Ouellet, B., & Roussel, M. (2001). About optimal timing and stability of weber fraction for duration discrimination. *Acoustical Science and Technology*, *22*(5), 370-372.
doi:10.1250/ast.22.370
- Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense". *Developmental Psychology*, *44*(5), 1457-1465. doi:10.1037/a0012682
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences*, *109*(28), 11116-11120. doi:10.1073/pnas.1200196109
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665-668.
doi:10.1038/nature07246
- Hillis, A. E., Rapp, B. C., & Caramazza, A. (1999). When a rose is a rose in speech but a tulip in writing. *Cortex*, *35*(3), 337-356. doi:10.1016/S0010-9452(08)70804-9
- Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17-29. doi:10.1016/j.jecp.2008.04.001
- Hope, J. (1989). Promoting number sense in school. *The Arithmetic Teacher*, *36*(6), 12.
- Howden, H. (1989). Number Sense. *Arithmetic Teacher*, *36*(6), 6-11.

- Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review*, *18*(6), 1222-1229. doi:10.3758/s13423-011-0154-1
- Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147-155. doi:10.1016/j.actpsy.2013.11.009
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, *106*(25), 10382-10385. doi:10.1073/pnas.0812142106
- Lemer, C., Dehaene, S., Spelke, E., & Cohen, L. (2003). Approximate quantities and exact number words: Dissociable systems. *Neuropsychologia*, *41*(14), 1942-1958. doi:10.1016/S0028-3932(03)00123-4
- Libertus, M. E. (2015). The role of intuitive approximation skills for school math abilities. *Mind, Brain, and Education*, *9*(2), 112-120. doi:10.1111/mbe.12072
- Libertus, M., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*(6), 1292-1300. doi:10.1111/j.1467-7687.2011.01080.x
- Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology*, *116*(4), 829-838. doi:10.1016/j.jecp.2013.08.003

- Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141*(3), 373-379.
doi:10.1016/j.actpsy.2012.09.009
- Libertus, M. E., Woldorff, M. G., & Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behavioral and Brain Functions, 3*(1), 1.
- Looi, C. Y., Thompson, J., Krause, B., & Cohen Kadosh, R. (2016). The neuroscience of mathematical cognition and learning. *Ministerio De Educacion*, Retrieved from <http://repositorio.minedu.gob.pe/handle/123456789/4665><http://repositorio.minedu.gob.pe/handle/123456789/4665>
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science, 17*(5), 714-726.
doi:10.1111/desc.12152
- Macaruso, P., McCloskey, M., & Aliminosa, D. (1993). The functional architecture of the cognitive numerical-processing system: Evidence from a patient with multiple impairments. *Cognitive Neuropsychology, 10*(4), 341-376. doi:10.1080/02643299308253468
- Maloney, E. A., Ansari, D., & Fugelsang, J. A. (2011). The effect of mathematics anxiety on the processing of numerical magnitude. *The Quarterly Journal of Experimental Psychology, 64*(1), 10-16. doi:10.1080/17470218.2010.533278
- Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*(4), 1224-1237. doi:10.1111/j.1467-8624.2011.01608.x

- Mazzocco, Feigenson, Halberda, & Santos. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, 6(9) Retrieved from <http://dx.plos.org/10.1371/journal.pone.0023749><http://dx.plos.org/10.1371/journal.pone.0023749>
- McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science*, 15(11), 776-781. doi:10.1111/j.0956-7976.2004.00755.x
- McIntosh, A., Reys, B. J., & Reys, R. E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics*, 12(3), 2-44. Retrieved from <http://www.jstor.org/stable/40248053><http://www.jstor.org/stable/40248053>
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9(3), 320-334. doi:10.1037/0097-7403.9.3.320
- Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 11(4), 591.
- Mussolin, C., Mejias, S., & Noël, M. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition*, 115(1), 10-25. doi:10.1016/j.cognition.2009.10.006

- Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2016). How approximate and exact number skills are related to each other across development: A review. *Developmental Review, 39*, 1-15. doi:10.1016/j.dr.2014.11.001
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.
- Ng, F. F., Tamis-LeMonda, C., Yoshikawa, H., & Sze, I. N. (2015). Inhibitory control in preschool predicts early math skills in first grade evidence from an ethnically diverse sample. *International Journal of Behavioral Development, 39*(2), 139-149. doi:10.1177/0165025414538558
- Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science, 297*(5587), 1708-1711.
- Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., & Content, A. (2013). Does math education modify the approximate number system? A comparison of schooled and unschooled adults. *Trends in Neuroscience and Education, 2*(1), 13-22. doi:10.1016/j.tine.2013.01.001
- Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General, 143*(1), 255.

- Odic, D., Lisboa, J. V., Eisinger, R., Olivera, M. G., Maiche, A., & Halberda, J. (2016). Approximate number and approximate time discrimination each correlate with school math abilities in young children. *Acta Psychologica, 163*, 17-26.
doi:10.1016/j.actpsy.2015.10.010
- Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*(1), 27-45.
- Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science, 42*(10). 0956797613482944.
doi:10.1177/0956797613482944
- Park, Y., & Cho, S. (2016). Developmental changes in the relationship between magnitude acuities and mathematical achievement in elementary school children. *Educational Psychology, 1*-15. doi:10.1080/01443410.2015.1127332
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences, 14*(12), 542-551. doi:10.1016/j.tics.2010.09.008
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*(1), 33-41. doi:10.1016/j.cognition.2010.03.012
- Piazza, M., & Izard, V. (2009). How humans count: Numerosity and the parietal cortex. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry, 15*(3), 261-273. doi:10.1177/1073858409333073

- Pinker, S. (1995). *The language instinct* (1st HarperPerennial ed.). New York: Harper Perennial.
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499-503.
- Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, *140*(1), 50-57.
doi:10.1016/j.actpsy.2012.02.008
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J., & De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, 1-16.
doi:10.1111/desc.12372
- Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PloS One*, *11*(4), e0153072.
- Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition*, *50*(1), 431-445.
- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: A critical review. *American Psychologist*, *60*(9), 950-958. doi:10.1037/0003-066X.60.9.950
- Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, *222*(4620), 179-181.
doi:10.1126/science.6623069

- Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, 36(2), 97-127. doi:10.1016/0010-0277(90)90001-Z
- Steen, L. A. (2001). Mathematics and numeracy: Two literacies, one language. *The Mathematics Educator*, 6(1), 10-16.
- Thompson, R. F., Mayers, K. S., Robertson, R. T., & Patterson, C. J. (1970). Number coding in association cortex of the cat. *Science*, 168(3928), 271-273.
- Wang, J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, 147, 82-99.
doi:10.1016/j.jecp.2016.03.002
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6389), 749-750.
doi:10.1038/358749a0
- Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, 89(1), 15.
- Xu, F., & Arriaga, R. I. (2007). Number discrimination in 10-month-old infants. *British Journal of Developmental Psychology*, 25(1), 103-108. doi:10.1348/026151005X90704
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74(1), B11. doi:10.1016/S0010-0277(99)00066-9

Crosscutting Concepts and Achievement: Is a Sense of Size and Scale Related to Achievement in Science and Mathematics?

Abstract

This study examined the relationship between students' (N= 229) concepts of size and scale and students' achievement in science and mathematics over a three-year period. Size and scale are considered a theme in science, which permeates disparate science and mathematics content areas, yet little is known about the relationship between students' conceptualization of size and scale and students' achievement in science and mathematics. Results indicated a strong positive significant relationship existed between students' understanding of size and scale and students' science achievement in grades five and eight. There was a positive significant relationship between students' concepts of size and scale and students' mathematics achievement in grades 5, 6, 7, and 8. Relationships between students' concepts of size and scale may be explained in part by the state curricula for each grade level. An examination of the relationships is included as well as an argument for the integration of crosscutting concepts into science and mathematics instruction as a way to support deep learning.

Crosscutting Concepts and Achievement: Is a Sense of Size and Scale Related to Achievement in Science and Mathematics?

National K-12 science standards have included an emphasis on broad themes in science as a way to unify seemingly disparate science content (American Association for the Advancement of Science [AAAS], 1994; National Research Council [NRC], 1996; 2012). As students construct meaning by making connections between prior knowledge and new experiences, broad themes provide a framework on which to base these connections, allowing for deep learning to occur. Standardized achievement tests are often used as a means of measuring students' comprehensive learning through a cumulative, standardized exam. However, there is a dearth of empirical investigations on the relationship between students' performance on standardized science and mathematics tests and students' concepts of crosscutting concepts.

One such crosscutting concept is scale, proportion, and quantity (NRC, 2012; Next Generation Science Standards [NGSS] Lead States, 2013) often referred to in the literature as *size* and *scale*. Size is defined as a descriptive or qualitative property such as bulk or magnitude (Magana, Brophy, & Bryan, 2012) while scale refers to "systems of measurement, which allow for the comparison of relative sizes" (Resnick, Davatzes, Newcombe, & Shipley, 2016, para. 1). Concepts of size and scale permeate scientific content. For example, in biology, the ratio of surface area to volume in cells dictates limitations in diffusion rates, thus restricting the size of cells. In physical science, molecular interactions at varying scales can lead to changes in the properties of materials. And in earth science, concepts of geologic time rely on the construction of appropriate scales that allow for the differentiation between

timescales that represent human history (e.g., hour, day, year, decade) and timescales that represent the history of all life on Earth (e.g., period, epoch, age). Despite the widespread presence across science content areas and claims that a strong understanding of crosscutting concepts is requisite for deep learning in science, the relationship between students' understanding of size and scale and students' achievement in science and mathematics standardized assessments has yet to be explored.

This investigation included a longitudinal panel study (Gall, Gall, & Borg, 2007) of students from sixth through eighth grade and addressed the following research questions:

- 1) *What is the relationship between students' concepts of relative scale and students' science and mathematics achievement in grades 6-8?*
- 2) *What is the relationship between students' concepts of absolute scale and students' mathematics achievement in grades 6-8?*

This research builds upon prior studies by exploring the claim that crosscutting concepts allow for deep learning of science content (NGSS Lead States, 2013). Unlike previous studies, that measured and reported only on students' concepts of size and scale over relatively short periods of time (e.g., Delgado, 2009; Jones & Taylor, 2009; Tretter, Jones, Andre, Negishi, & Minogue, 2006), this investigation provided empirical evidence linking broad themes to performance on standardized tests over a three year period.

In an effort to establish the relevance of size and scale as it relates to general ability in science, this study examined the relationship between students' concepts of size and scale and students' achievement on standardized science and mathematics tests. If a positive relationship exists between understanding scale and achievement in science and mathematics

as measured by standardized tests, it can provide evidence for the premise that crosscutting concepts can promote general understanding in science and mathematics. Uncovering this relationship may not shed light on the degree to which scale contributes to achievement nor will it indicate if concepts of size and scale are a component of another skillset or construct such as multiplicative reasoning, spatial visualization, or general ability in mathematics or science. However, understanding the relationship between students' concepts of size and scale and students' achievement can provide insight into whether or not an emphasis on crosscutting concepts in science instruction is warranted. If there is no relationship between students' concepts of size and scale and students' achievement then new questions arise about the developmental appropriateness of size and scale as a broad theme as well as the validity of the standardized tests as measures of scientific knowledge. Thus, students could have a strong sense of scale but be unable to recall vocabulary or specific content assessed by the standardized test. Another explanation is that if there is no relationship between students' achievement and students' concepts of size and scale, then this crosscutting concept may not contribute to overall understanding of science and mathematics.

Literature Review

Three Dimensional Learning. In an effort to shift science instructional practices, the National Research Council outlined a three dimensional pedagogical approach in which students engage in deep learning of science and engineering through the integration of disciplinary core ideas, science and engineering practices, and crosscutting concepts (NRC, 2012). Krajcik, Codere, Dahsah, Bayer, and Mun (2014) explained that like strands in a rope, the three dimensions work together to “build an integrated understanding of a rich network of

connected ideas. The more connections developed, the greater the ability of students to solve problems, make decisions, explain phenomena, and make sense of new information,” (p. 157). This three dimensional pedagogy served as the foundation for developing the *Next Generation Science Standards* (NGSS), a list of grade-level standards that present science standards as performance expectations, rather than isolated facts (NGSS Lead States, 2013). The performance expectations were designed so that students would have to use disciplinary core ideas, science and engineering practices, and crosscutting concepts to explain scientific phenomena, which would encourage deep learning in science (NRC, 2012). Underpinning the motivation for deep learning in science is the need for a scientifically literate citizenry. While the definition of science literacy takes many forms, scholars broadly agree that it refers to the familiarity with science held by the public (Deboer, 2000) and the application of science to everyday experiences (Bybee, 2015). A goal of science education is to create scientifically literate citizens that are capable of making science-based or evidence-based decisions, necessary skills in the 21st century. There is growing focus on research and dialogue related to the NGSS as a whole (e.g., Bybee, 2014; Krajcik et al., 2014; Stage, Asturias, Cheuk, Daro, 2013), but less is known about crosscutting concepts, such as size and scale, and the relationship between conceptual understanding of the crosscutting theme and science learning.

Crosscutting Concepts. The National K-12 science standards have highlighted the need for integrating cross-curricular concepts in science as a means of unifying science content (AAAS, 1994; NRC, 1996; 2012; NGSS Lead States, 2013). One such crosscutting theme that may promote both deep learning and scientific literacy, is scale, proportion, and

quantity (NRC, 2012; NGSS Lead States, 2013), referred to in this study as *size* and *scale*. Magana et al. (2012) differentiated between size and scale by defining size as a descriptive quality such as bulk or magnitude and defining scale as a quantifiable characteristic that can incrementally increase or decrease in measurable units. These measurable units help form measurement systems that allow for comparisons to be made between relative quantities. For example, measurable units such as hours and centimeters each provide a scale that allows for the comparison of quantities of time (temporal duration) and space, respectively (Resnick et al., 2016). The NGSS posited that concepts of size and scale are necessary for deep learning of scientific content because “in considering phenomena, it is critical to recognize what is relevant at different measures of size, time, and energy and to recognize how changes in scale, proportion, and quantity affect a system’s structure or performance,” (NGSS Lead States, 2013, p. 84). While specific science content can be taught across science domains, it is generally accepted that a conceptual understanding of scale, proportion, and quantity is a prerequisite for understanding science concepts such as cells and cell size, evolutionary or geologic time, and atomic interactions yet, at this time, there are no empirical studies investigating the relationship between students’ understanding of size and scale and students’ achievement in science and mathematics.

Measurement and Estimation. Fundamental to concepts of size and scale is numerosity or quantity. When young children first learn quantity, they conceptualize it through making connections between physical objects and representative numerical symbols (Booth & Siegler, 2006; Carey, 2004; Joram, Subrahmanyam, & Gelman, 1998). One way to demonstrate the relationship between quantity and space is through measurement activities.

Not surprisingly, measurement is often used as a means of determining students' concepts of quantity. In studying students' understanding of part to whole, partitioning, and conservation of length using physical objects, researchers highlighted evidence of a developmental component to understanding the size of objects and the relationship between objects and their surroundings (Piaget, Inhelder, & Szeminska, 1960; Piaget & Inhelder, 2013). More recently, Lehrer (2003) found that students' understanding of units, iteration, tiling, standardization, proportionality, additive properties, and origin of location all laid the foundation for understanding measurement and units. In a review of measurement and estimation research, Joram et al. (1998) outlined three paradigmatic shifts in measurement and estimation research. The first of which sought to answer the question of how well children and adults could estimate. While the reviewers found that most research reported individuals were poor estimators, Joram et al. also reported that that accuracy in estimation increased with age. The early studies were not without flaws; initial research designs involving estimation had participants guess the sizes of objects that were not physically present, resulting in possible subjectivity among participants' estimations (Joram et al., 1998). The second generation or shift in measurement and estimation research during the 1960's and 1970's expanded measurement beyond linearity into other attributes such as weight, temperature, or volume, and the objects being measured were physically present (Joram et al., 1998). Findings indicated that individuals were the most accurate in measuring temperature, followed by linear measurements, but other estimations were poor (Joram et al., 1998). During the 1980's there was another paradigmatic shift in measurement and estimation research, marked by the movement from accuracy in measurement estimation to the cognitive process involved in

estimation (Joram et al., 1998). This research movement was criticized for its disconnect from instruction in measurement, inadequacy in describing the cognitive process, and inclusion of cumbersome models that failed to distinguish between discrete and continuous quantities (Joram et al., 1998).

Building on prior research in estimation, Albarracín and Gorgorió (2014) more recently defined estimation as the as a “rough calculation or judgment of the value, number, quantity, or extent of something,” (p. 81) and explained further that estimation is useful because estimated values provide insight into individuals’ perceptions and judgments about the estimated item. Hogan and Brezinski (2003) distinguished between three areas of estimation: numerosity, estimation of magnitudes or measurements, and computational estimation.

Numerosity refers to the ability to estimate the number of objects in an array...

Measurement estimation requires the participant to provide estimates of length, height, weight, liquid capacity, and similar measures, usually of common objects in the environment... Computational estimation refers to providing estimated answers to [mathematical] computations... (p. 260).

Albarracín and Gorgorió (2014) built on these three types of estimation and added another type of estimation, the use of models representing real events or items. In a study of students aged 12 to 16, Albarracín and Gorgorió found students employed iteration of units, concentration measurements, and grid distributions while solving Fermi problems (e.g., Efthimiou & Llewellyn, 2007) using large numbers. The use of modeling to demonstrate the contextualization of measurement estimation and numerosity has been employed by science

education researchers in understanding students' concepts of size and scale as it relates to specific science contexts ranging from the atomic scale to deep time (e.g., Delgado, 2014; Jones, Forrester, Robertson, Gardner, & Taylor, 2012; Jones, Gardner, Taylor, Wiebe, & Forrester, 2011; Joram et al., 1998; Trend, 2001; Tretter, Jones, Andre et al., 2006). The following section summarizes the literature on students' concepts of size and scale in the specific context of science.

Size and Scale in Science Education. Similar to early work in estimation, preliminary science education research in size and scale involved understanding concepts of linear scale, specifically as it related to navigation and mapping (e.g., Golledge, Gale, Pellegrino, & Doherty, 1992). More recently, research in geography and way finding has shifted to focus on spatial cognition with some emphasis on spatial thinking in science (Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Hegarty, 2014; Liben, Kastens, & Christensen, 2011; Liben, Myers, Christensen, & Bower, 2013; Vandelacluze, 2012). Spatial ability differs from concepts of scale in that it refers to how an individual might imagine inspecting, transforming, or manipulating an object in space (e.g., rotation), actions which can occur at any scale (Hegarty et al., 2006). Hogan and Brezinski (2003) found that numerosity and measurement estimation ability were correlated to spatial ability in college students. Though related, measurement estimation ability differs from spatial ability in that measurement estimation emphasizes the iteration of units to determine the magnitude or size of an object, while spatial ability emphasizes the orientation of the object in space at any given size. Hogan and Brezinski argued that while both numerosity and estimation of magnitudes relied on spatial and perceptive abilities, measurement estimation ability failed to

follow general mathematical ability and, therefore, it should be considered a separate skill. While spatial abilities are strong predictors of science achievement and science attainment (Shea, Lubinski, & Benbow, 2001; Wai, Lubinski, & Benbow, 2009) it is still unclear the role that concepts of size and scale, as indexed by measurement estimation ability, play in science and mathematics achievement.

Concepts of geologic time are another area of interest to science education researchers. Studies of individuals' concepts of geologic or deep time are twofold in that they include both the succession and duration of geologic events (Delgado, 2014). Studies on succession indicated that individuals have a more detailed understanding of events that occurred closer to human existence than those from earlier time periods and that individuals categorize time into three periods: extremely ancient, moderately ancient, and less ancient (Trend, 2001). Delgado (2014) used the three aforementioned categories to evaluate undergraduate students' ability to estimate time scales related to landmarks (e.g., major cosmological, geological, or historical events). Similar to Trend (2001), Delgado found that students had greater accuracy identifying landmarks at temporal extremes (e.g., big bang, human civilization), with the most accurate landmarks identified within human history. Cheek (2013) focused on undergraduate students' use of geologic landmarks to explore the students' views of the relationship between spatial size and duration of time, and found that even as young adults, individuals struggled with concepts of size and scale. Through task-based interviews, Cheek found that students (N=17) were inclined to equate spatial size with temporal duration (i.e., the length of a timeline represented a model of geologic time) but were often confused by several key factors. Students had difficulty understanding time

periods associated with large numbers (e.g., time periods of up to 100,000,000 years). Students often skewed spatial size such that there was a mismatch between temporal representation and spatial size (i.e., long time periods were too short, short periods were too long). Lastly, students often demonstrated a lack of use of events or markers to denote temporal duration (Cheek, 2013). These aforementioned studies highlighted both the difficulty individuals have with understanding size or quantity (e.g., age of the Earth in years) and scale (e.g., spatial and temporal mismatch). The studies also provided some insight into how students mentally maneuver across orders of magnitude through the interpretation of temporal and spatial relationships (e.g., Cheek, 2013) as well as the use of landmarks (e.g., Delgado, 2014).

Tretter, Jones, Andre, et al. (2006) conducted one of the earliest studies on individuals' concepts of size and scale. Tretter and colleagues found that all participants (N=215), from doctoral students to grades 5, 7, 9, 12 were more accurate in generating relative sizes of objects than they were with actual sizes of objects. Tretter, Jones, Andre et al. argued that this was likely due to the more frequent use of relative sizes in everyday life. In contrast to the lack of landmarks used by participants in reasoning about geologic time (Cheek, 2013), Tretter and colleagues found that participants often anchored relative sizes of objects to landmarks. The authors described these landmarks as objects or distances with which participants had personal experiences and found that the number of landmark categories increased with expertise (Tretter, Jones, Andre et al., 2006). For example, elementary students grouped objects ranging in size from the nanoscale to the galactic scale into only four size landmark categories (e.g., big, field size, room size, and small). In

contrast, doctoral students grouped the same list of objects into ten landmark categories (e.g., big, travel size, field size, room size, me, small, very small, barely visible, many atoms, and atomic). Tretter, Jones, Andre et al. also found that accuracy in relative sizes and absolute sizes improved with both age and expertise.

As part of the same study, Tretter, Jones, and Minogue (2006) investigated the difference in accuracy of conceptions between large scale and small scale. Similar to findings from the research in individuals' conceptualization of geologic time (e.g., Delgado, 2014; Trend, 2001), Tretter, et al. found participants' accuracy decreased as objects under review moved away from human-sized scales. Different from the geologic time scales, however, was the way in which individuals conceptualized extremes. For example, Tretter, Jones, and Minogue found asymmetrical conceptions of large and small scales, with large-scale accuracy decreasing at fairly even intervals with increasing scale but small-scale accuracy failing to demonstrate a similar pattern. This led the authors to conclude that concepts of scale are less accurate as they move away from human sizes, with smaller scales being more difficult for participants to conceptualize than larger scales (Tretter, Jones, & Minogue, 2006). In addition, these researchers found that experts leveraged mathematics when mentally maneuvering between different scaled worlds (i.e., using exponents or powers of ten when thinking about different scales) and that these transitions between worlds were not smooth, but rather occurred in leaps or jumps from one scale to the next (Tretter, Jones, & Minogue, 2006).

Building on this idea of mentally maneuvering through leaps with mathematics, Jones et al. (2007) evaluated the efficacy of the film *Powers of Ten*, from the perspectives of both

teachers and students. Results indicated that the film improved students' concepts of scale as well as understanding of metric sizes and scientific notation. Similar to previous studies (e.g., Trend, 2001; Tretter, Jones, Andre et al., 2006; Tretter, Jones & Minogue, 2006), students reported difficulty with conceptualizing sizes and scales outside of the human experience, with smaller extremes more difficult to understand than larger extremes (Jones, Taylor, et al., 2007). Additionally, experienced teachers reported that the pace at which the film slowly navigated from human scale to larger or smaller scales and then more quickly navigated back to the human scale appropriately scaffolded the concepts of size and scale for students (Jones, Taylor, et al., 2007).

Jones and Rua (2008) investigated conceptual representations of flu and microbial illness held by students, teachers, and medical professionals and found that characteristics about the illnesses were better known across groups than the size and scale of the disease-causing agents. Not surprisingly, students' experiences and conceptions about the microbial illness influenced students' understanding of the size of the microbe. For example, elementary and middle school students shared the idea that viruses and bacteria differed in size due to differences in virulence. Students explained viruses were larger than bacteria because they perceived viruses as something that could make them sicker than bacteria (Jones & Rua, 2008).

Similar to the above comparison of medical professionals, teachers, and students, Jones et al. (2008) compared expert and novice teachers' concepts of spatial scale. Not surprisingly, both groups demonstrated the greatest degree of accuracy at the human scale, with larger scales being more accurate than smaller scales (Jones et al., 2008). In support of

previous findings, those individuals with more experience teaching were more accurate at smaller scales (Jones et al., 2008). Furthermore, personal experiences with the objects and distances affected both groups' understanding of size and scale (Jones et al., 2008).

Building on the idea of the influence of personal experiences with size and scale and the ability of individuals at a wide range of ages and levels of expertise to comprehend scale at the human size, Jones, Taylor, and Broadwell (2009) investigated the efficacy of using metric body measures to teach students about estimation and measurement. Jones et al. found that using a rough body measure as a tool significantly improved students' (N=19) accuracy in estimating metric lengths. Researchers also found a relationship between proportional reasoning and performance on measurement assessments (Jones et al., 2009). The results of this investigation led the same researchers to take a deeper look into the influence of proportional reasoning on students' concepts of size and scale.

One key application of size and scale to science content is the concept of the relationship between surface area and volume within a system. Taylor and Jones (2009) found that proportional reasoning in middle school participants (N=19) was significantly correlated with students' concepts of surface area and volume. Proportional reasoning may play a role in the ways in which students conceptualize fractions, particularly those representative of very small numbers, and may explain the lack of understanding of small scales in previous studies (Jones et al., 2008; Tretter, Jones, Andre, et al., 2006; Tretter, Jones & Minogue, 2006).

Jones and Taylor (2009) investigated the types of scales and scaling strategies used by professionals in different work domains, the in and out of school experiences that

contributed to those strategies and perceived skills, and how professionals perceived the role of scale in their work. Similar to findings by Tretter, Jones, Andre et al. (2006), professionals made use of anchoring while maneuvering across scales at work and emphasized the importance of scale in their professional work (Jones & Taylor, 2009). More than half of the participants (N=50) reported that their sense of scale likely developed through kinesthetic or physical experiences (Jones & Taylor, 2009).

Recent research has focused not only an individual's concepts of size and scale, but also the cognitive processes and reasoning associated with an understanding of size and scale. For example, Resnick et al. (2016) argued that individuals reason about things outside the human-scale experience with support from relational reasoning. Relational reasoning refers to a cognitive mechanism, which uses comparisons (both similarities and differences) to construct conceptual categories that allow individuals to make meaning of their world (Resnick et al., 2016). Representational similarities include analogies that allow individuals to map or recycle the human-scale onto extreme-scales. For example, if the length of the human body represents the age of the Earth, the very tip of one fingernail would represent humans' evolutionary existence on Earth. This analogy allows students to recycle small number processes with which they are familiar (e.g., size of the human body) to support reasoning about an extreme scale with which they may not be familiar (e.g., age of the Earth). According to Resnick et al., dissimilarities also support understanding scale because they allow individuals to observe differences that exist at different scales. For example, changes in the temperature of the Earth have very different effects at different temporal scales. An increase of five degrees Fahrenheit may have negligible effects over the course of

a single day but much different effects over the age of the Earth (Resnick et al., 2016). Resnick et al. argued that dissimilarities, in particular, allow for conceptual change as it relates to reasoning about scale. While Resnick and colleagues have found this framework useful in designing successful interventions in teaching about scale, more research is needed for understanding cognitive processes associated with size and scale, best practices for incorporating size and scale into everyday instruction, and the effects of these interventions on students' science and mathematics achievement. Furthermore, the dissimilarities outlined by Resnick et al. are strikingly similar to the components of conceptual change theory, specifically, the idea of *dissatisfaction*, (Strike & Posner, 1992) and seem to present a well-established framework in a specific context under a new name.

Methodology

Participants. The population sample included a group of sixth grade students at an urban public magnet middle school in the southeastern United States. Permission forms were sent home to all students and those whose parents provided consent were provided the opportunity to participate in the study. The school was selected as a convenience sample because it was in close proximity to the research site and there were existing partnerships between the school and the university. Forty-three percent of all students (grades 6-8) at the school were proficient on all End-of-Grade (EOG) achievement tests in 2015, as compared to 66.1% and 56.3% proficiency measured at the district level and state level, respectively (NC Report Card, 2015).

Longitudinal Design. The study adopted a modified panel longitudinal research design, as it followed the same class of students over a three-year period. Gall et al. (2007)

defined panel longitudinal research as “a type of investigation in which changes in a population over time are studied by selecting a sample at the outset of the study and then collecting data from the same sample throughout the duration of the study” (p. 631). This form of research design differs from both cohort studies and trend studies in that the sample of participants does not change at each data collection point. This study slightly modified the panel longitudinal design because some students were unable to participate in each year of the study, though efforts were made to retain as many students as possible from year to year. In grade eight, 109 additional students were added to the panel of existing participants to increase the sample size and provide a more complete picture of the relationship between students’ concepts of scale and students’ academic achievement at the school. The three-year design of the study provided an opportunity to see how students’ understanding of an abstract concept such as scale changed across various stages of cognitive development, and a panel design was particularly useful for comparing the long-term effects of the relationship between students’ understanding of size and scale and students’ achievement in science and mathematics across multiple years.

Longitudinal research is not commonly done due to the difficulties in tracking participants from one year to the next. Simpson and Oliver (1990) examined literature on longitudinal studies involving science achievement and included both decade-long studies (e.g., Coleman, 1981) as well as studies spanning only two years (e.g., Handley & Morse, 1984) in their review and found that most studies included large sample sizes ($N > 1,000$) that followed students for about a decade. Table 1 includes a description of the sample size in each of the three years of the study.

Table 1

Participant sample size for each assessment across all grades

<u>Assessments</u>	<u>Grade</u>			
	<u>Pre-6</u>	<u>Post-6</u>	<u>Post-7</u>	<u>Post-8</u>
Relative Scale (SCS)	114	117	115	214
Absolute Scale (SOQ)	179	125	120	229

Assessments. To address the first research question and determine the relationship between students' concepts of absolute size and students' achievement in science and mathematics, students completed the Scale of Objects Questionnaire (SOQ) (see Tretter, Jones, Andre et al., 2006; Tretter, Jones & Minogue, 2006). Efforts were made to ensure that the assessments used in the study to measure students' concepts of size and scale had achieved adequate validity and reliability, as these instruments have been used in a number of studies (e.g., Jones, Taylor, Minogue, Broadwell, Wiebe, & Carter, 2007; Jones, Tretter, Taylor, & Oppewal, 2008; Tretter, Jones, Andre, et al., 2006; Tretter, Jones, & Minogue, 2006).

Tretter, Jones, Andre et al. (2006) developed the 26-question SOQ based on work by Trend (2001). Tretter, Jones, Andre et al. piloted the SOQ with both students and teachers, made necessary revisions, piloted the questionnaire again, and made final revisions prior to using it in their research. The SOQ has been employed to index students' concepts of absolute scale in a number of publications (e.g., Tretter, Jones, Andre, et al., 2006; Tretter,

Jones, & Minogue, 2006). The SOQ asked students to determine if the size of an object falls into one of twelve size categories. For example, students were asked to determine the size category for the width of a soccer ball by indicating that the ball was between one tenth of a meter and one meter. The internal reliability of the SOQ was established by a Cronbach's' Alpha of 0.763 for the sample population. To address research question two and determine students' concepts of relative size, students completed the Scale Card Sort (SCS). The SCS was based on the sorting task developed by Tretter, Jones, Andre, et al. (2006) and later modified by Jones, Taylor, et al. (2007) to index students' concepts of relative sizes and distances. The SCS included 31 cards that students sorted by size from smallest item to largest item (see Appendix A). During the administration of this instrument, students divided all the randomly assorted cards into two groups: large items and small items. Students worked with only one of the groups (large or small) at a time and then later combined the two piles. This scaffolded the assessment for students by decreasing the number of physical objects students manipulated at one time. Pilot testing demonstrated that this approach benefitted 6th grade students' successful completion of the SCS task. The internal reliability of the SCS was established by a Cronbach's' Alpha of 0.953 for the sample population.

Experimental Design. This investigation sought to determine the relationship (if any) between students' concepts of size and scale and students' achievement in science and mathematics. Data were collected from grade 5 to grade 8. The scale assessments (SOQ and SCS) were administered pre and post in grade 6 and post only in grades 7 and 8. The mathematics EOG was administered as a post-test in grades 5 through 8 and the science EOG was administered as a post test in grades 5 and 8. The research team agreed that a negligible

amount of instruction likely took place between post assessments in grade 5 given at the end of the school year and pre assessments in grade 6 given at the beginning of the school year. For that reason, student performances on grade 5 EOGs were considered temporal equivalents to the grade 6 pre-assessments on SOQ and SCS. Table 2 outlines the administration of each of the assessments across the course of the study.

Table 2

Administration of each data collection instrument during the multi-year study.

<u>Instruments</u>	Grade							
	5		6		7		8	
	<u>Pre</u>	<u>Post</u>	<u>Pre</u>	<u>Post</u>	<u>Pre</u>	<u>Post</u>	<u>Pre</u>	<u>Post</u>
SOQ			x	x		x		x
SCS			x	x		x		x
EOG Mathematics		x		x		x		x
EOG Science		x						x

Analyses.

Scoring Assessments. Scores for both SOQ and SCS were calculated based on the number of correct and nearly correct (within a power of ten) answers. On the SOQ, students earned one point for selecting the correct size category and earned half of a point for selecting a size category that was within one power of ten of the correct answer. Similarly, on the SCS, students earned one point for placing a card in the correct sequence (i.e., the second

smallest item in cards was the second card in the submitted sequence of sorted cards) and students earned half a point for placing a card within one position from the correct sequence (i.e., the second smallest item in the cards was the third card in the submitted sequence of sorted cards making it within one position from the correct sequence).

Science and mathematics achievement data were collected from developmental scale scores on the science and mathematics EOGs. On all EOG assessments, the number of correct responses on the tests determines students' raw scores and the raw score is converted into a developmental scale score (North Carolina Department of Public Instruction [NCDPI], 2014). The mathematics EOGs were administered in grades 5 through 8. The fifth grade mathematics EOG consisted of 46 multiple-choice items and 8 gridded response items. The mathematics EOG in grades 6-8 consisted of 60 total items, 49 of which were multiple choice and 11 of which were gridded response (NCDPI, 2014.) The science EOGs were administered in grade 5 and grade 8, and both instruments consisted of 75 multiple-choice items (NCDPI, 2014). The EOGs are both valid and reliable standardized assessments developed by the North Carolina Department of Public Instruction and are used, in part, by the school system to determine if students have acquired the knowledge necessary to advance to the next grade level.

Preliminary Analyses. Independent sample t-tests indicated that there were no statistically significant differences in mean scores on any of the assessments between students who completed less than half of all 14 assessments and students who completed at least half of all 14 assessments (see Appendix B). This serves as an indication that students with missing values or incomplete values should be included in the overall analyses. The

goal of this study was to explore whether or not students who had high achievement in science and mathematics also had strong conceptual understandings of size and scale. For this reason, correlational analyses were used and are discussed in the following section.

Correlation Analyses. Correlational analyses using Pearson's correlation coefficient were used to address both research questions and determine the relationship between students' concepts of scale and students' achievement in science and mathematics. Analyses were conducted for each assessment in each year and are reported in aggregate for each of the twelve relationships, and p-values less than 0.05 were used to indicate significance. This approach allowed for the inclusion of all students with data for at least one scale assessment (SOQ or SCS) and at least one EOG (mathematics or science). Students for whom there was not enough data to report on at least one relationship among the twelve combinations of variables, were not included in the study.

Pearson's product-moment correlations were run to evaluate each of the twelve relationships. The null hypothesis for this test is that the correlation coefficient (R) for each relationship examined equals zero. The alternative hypothesis is that R is not equal to zero. Rejecting the null hypothesis would mean that there is a statistically significant relationship between the two factors in each of the correlations. Preliminary analyses showed that each relationship was linear. Preliminary analyses also showed that not all variables were normally distributed as assessed by Shapiro-Wilk's test ($p < 0.05$), but because Pearson's correlation is somewhat robust to deviations from normality, the variables remained in the analyses. A visual inspection of scatterplots of each relationship prompted the removal of outliers in grades 5 and 6 of no more than 8 total cases (i.e., students) in each correlation. No

outliers were identified or removed in grades 7 and 8. A Fisher's r-to-z transformation (Cohen & Cohen, 1983) was used to determine if there were significant differences between correlations. Correlation coefficients were converted into z-scores which, coupled with sample sizes, were compared to determine if there were statistically significant differences in the correlations (Preacher, 2002).

Limitations of the Study. There are several limitations to this study that should be acknowledged. The first is that the sample of students may not be a representative sample of all students in middle grades, though efforts were made to include public school students since those that attend public schools tend to be from a variety of backgrounds related to racial, ethnic, socioeconomic, and academic factors. Another limitation to the study is that students' who participated in all three years of the assessments may have experienced some degree of assessment fatigue, which may have discouraged their engagement or participation in later years of the study. Participant attrition in the study also served as a limitation, as some students moved away before completion of the project. To address the declining number of participants in grade 8, the assessments were administered to all students in grade 8 to maintain an adequate sample size.

Results

The results of this study indicated that there was a significant relationship between students' concepts of scale and students' achievement in science and mathematics for all years in the study (see Table 3). Students' performance on absolute scale instruments demonstrated a stronger, positive correlation with achievement than students' performance on relative scale instruments (see Table 3). Furthermore, science achievement demonstrated a

stronger relationship with concepts of scale than mathematics achievement, though this difference was not significant and is discussed in more detail below. The following results are organized by subject area, with science achievement followed by mathematics achievement.

Table 3

Pearson's correlation coefficients for the relationship between achievement and concepts of scale.

	Science Achievement		Mathematics Achievement			
	<u>Post-5/Pre-6</u>	<u>Post-8</u>	<u>Post-5/Pre-6</u>	<u>Post-6</u>	<u>Post-7</u>	<u>Post-8</u>
Relative Scale (SCS)	0.48**	0.55**	0.35**	0.56**	0.51**	0.52**
Absolute Scale (SOQ)	0.73**	0.73**	0.66**	0.25**	0.61**	0.68**

Note. SCS= Scale Card Sort, SOQ = Scale of Objects Questionnaire, ** = statistically significant at $p < 0.01$ level.

Science. Science achievement was collected in grades 5 and 8 by the school system. Students' science achievement and students' concepts of relative and absolute scale were positively and significantly correlated for grades 5 and 8, on both relative and absolute scale assessments. Table 3 lists Pearson's correlation coefficients (R) and significance levels for each of the relationships examined. The relationship between students' concepts of absolute scale and students' science achievement was significantly stronger than the relationship

between students' concepts of relative scale and students' science achievement in grade 5, $Z = -3.37, p < 0.05$ as well as grade 8, $Z = -3.19, p < 0.05$. Although the relationships between concepts of scale (both relative and absolute) and science achievement were stronger than the relationships between concepts of scale and mathematics achievement, this difference was not significant for both relative scale in grades 5 ($Z = 1.15, p = 0.125$) and 8 ($Z = 0.36, p = 0.36$) as well as absolute scale in grades 5 ($Z = 1.36, p = 0.09$) and 8 ($Z = 0.95, p = 0.17$).

Mathematics. Mathematics achievement was collected in grades 5 through 8 by the school system, and findings indicated that a significant, positive correlation existed for all grades (see Table 3) between students' achievement in mathematics and students' concepts both relative and absolute scale. In general, the strength of the relationship increased with increasing grade level for both relative and absolute scale assessments. Similar to the trends shown with science achievement, mathematics achievement, demonstrated a stronger relationship with absolute scale than relative scale in grades 5/Pre-6 ($Z = -3.47, p < 0.05$), 7 ($Z = -1.15, p = 0.13$), and 8 ($Z = -2.62, p < 0.05$), though the difference in correlations was not significant for grade 7. However, this same trend did not hold for grade 6, with relative scale demonstrating a significantly stronger correlation to mathematics achievement than absolute scale, $Z = 2.88, p < 0.05$.

Discussion

The results of this investigation bring to light the relationship between students' concepts of size and scale and students' academic achievement. Due to the limitations of the correlational analyses, the results of this study are unable to support the idea that concepts of scale are requisite for deep learning in science; however, it does present a preliminary

argument for enhanced investigations of the role of crosscutting concepts in science and mathematics education. The following sections outline possible explanations for the findings of the study, directions for future research, and implications for science teaching.

Contextualization of Size and Scale within State Standards. Although the standardized tests in science for this state are based on the state standards and not the Performance Expectations outlined in the NGSS, the crosscutting concept of scale, proportion, and quantity, permeates much of the state standards (see Table 4). One explanation of the relationship observed between students' concepts of size and scale and students' academic achievement is that the standardized assessments do, in fact, measure students' understandings of crosscutting concepts. The standardized assessments measure students' general science and mathematics ability and the correlation observed may reflect that students with generally high science and mathematics achievement also have strong concepts of size and scale. The following sections discuss the observed relationships between students' concepts of size and scale and students' academic achievement within the context of the state standards for science and mathematics.

Science Standards. The relationship between students' achievement in science and students' concepts of scale did not vary from grade 5 to grade 8, indicating that there was a significant, positive correlation between students with high science achievement and students with strong concepts of size and scale in both grades. This similarity across grade level may be the result of classroom instruction, which likely aligns with state standards in each grade. The grade 5 standards could be taught in a way that contextualized the core content within the crosscutting concept of scale, proportion, and quantity because the scientific phenomena

they address, are presented at the human scale but could be explained beyond the human scale. For example, the standard on weather recommends students understand weather patterns and “making connections to the weather in a particular place and time,” failing to suggest students learn about weather within the context of climate or global climate changes (NCDPI, n.d.-a, p. 8). Yet, it is possible that teachers explained weather within the context of climate, which could situate weather patterns within geologic time, bringing in sizes and scales outside of the human scale, thus helping to explain why students who performed well on the size and scale assessments also had high achievement in science. Another example of content that could be addressed within the context of the crosscutting concept of size and scale is the grade 5 standard on inheritance. The inheritance science standard includes an examination of inheritance patterns at the observable, human scale. Though not mentioned in the standards, it is likely that teachers may have included the role that compounds at the micro or nanoscale (e.g., organelles, proteins, DNA) play in determining macroscale observations such as human traits (NCDPI, n.d.-a). This could play a pivotal role in developing students’ primary concepts of scale, as the introduction to the existence of the unseen world (i.e., micro or nanoscale) has been noted in learning trajectories as an early developmental component to understanding size and scale (Delgado, 2009; Jones & Taylor, 2009). Teachers in the study likely tailored their instruction to meet the state standards to prepare their students for the state summative assessment. The grade 5 science standards and assessments may be written in such a way that students who have a strong understanding of size and scale are also able to perform well on tests that measure the state standards, as

reflected in the relationship between students' concepts of size and scale and students' science achievement.

Compared to the grade 5 state standards, the grade 8 state standards more explicitly incorporate different sizes and scales within the context of science (NCDPI, n.d.-b). For example, rather than one particular weather incident, as depicted in grade 5, the grade 8 standard about weather required students examine the hydrosphere as a whole, a larger and more abstract size, as it includes all the water on Earth rather than one singular weather event (NCDPI, n.d.-b). In addition, the history of life on earth and the evolution of species cannot be taught apart from the understanding of the age of the Earth and the abstract numbers associated with that age (i.e., million years or billion years) (NCDPI, n.d.-b). For this reason, to address the state standards, teachers in grade 8 may have made efforts to explicitly contextualize content such as the history of life on Earth by teaching about size and scale. Table 4 lists the state standards in science for both grade 5 and grade 8. This relationship supports the supposition that students' learning of science content is enhanced through the incorporation of broad themes. Though standardized tests have undoubtedly changed the landscape of public of education, the results of this study indicate that a relationship exists between students' performance on standardized tests for mathematics and science and students' concepts of size and scale. Such a relationship may suggest that developing a sense of scale parallels conceptual understanding of science or that conceptual understanding of science parallels the development of a sense of scale. Another explanation could be that a reciprocal relationship exists between understanding scale and understanding science content, as one cannot be adequately explained without the other. Further research is needed to

examine the causal relationship between concepts of scale and student achievement in science.

Table 4

Descriptions of the State Essential Standards for science in grades 5 and 8 (NCDPI, n.d.-a; n.d.-b)

State Essential Standards for Science	
<u>Grade 5</u>	<u>Grade 8</u>
5.P.1 Understand force, motion and the relationship between them.	8.P.1 Understand the properties of matter and changes that occur when matter interacts in an open and closed container.
5.P.2 Understand the interactions of matter and energy and the changes that occur.	8.P.2 Explain the environmental implications associated with the various methods of obtaining, managing, and using energy resources.
5.P.3 Explain how the properties of some materials change as a result of heating and cooling.	8.E.1 Understand the hydrosphere and the impact of humans on local systems and the effects of the hydrosphere on humans.
5.E.1 Understand weather patterns and phenomena, making connections to the weather in a particular place and time.	8.E.2 Understand the history of Earth and its life forms based on evidence of change recorded in fossil records and landforms.

Table 4 Continued

State Essential Standards for Science

<u>Grade 5</u>	<u>Grade 8</u>
5.L.1 Understand how structures and systems of organisms perform functions necessary for life.	8.L.1 Understand the hazards caused by agents of diseases that affect living organisms.
5.L.2 Understand the interdependence of plants and animals with their ecosystem.	8.L.2 Understand how biotechnology is used to affect living organisms.
5.L.3 Understand why organisms differ from or are similar to their parents based on the characteristics of the organism.	8.L.3 Understand how organisms interact with and respond to the biotic and abiotic components of their environment.
	8.L.4 Understand the evolution of organisms and landforms based on evidence, theories and processes that impact the Earth over time.
	8.L.5 Understand the composition of various substances as it relates to their ability to serve as a source of energy and building materials for growth and repair of organisms.

Mathematics Standards. Because a positive and significant relationship between mathematics and scale existed for all grades, one argument is that mathematics is the underlying factor that consistently explains students' concepts of scale over time. Another explanation of the positive and significant correlation between concepts of scale and mathematics achievement that existed in all grades may be a result of the emphasis of scaling in the state standards in mathematics, the document on which the standardized achievement test in mathematics is based. In grade 5, there was a stronger correlation with mathematics achievement on the absolute scale instrument than the relative scale instrument. This may be explained by the emphasis in the grade 5 standards on topics such as place value, fractions, and measurement conversions (NCDPI, 2015), all of which are closely tied to concepts of absolute scale. To address these mathematics standards, teachers in grade 5 may have emphasized mathematics content that supports understanding of scale such that students with high mathematics achievement also had high conceptual understandings of size and scale. In contrast, results in grade 6 indicated that the relationship between relative scale and mathematics achievement was stronger than the relationship between absolute scale and mathematics achievement. This may be explained by the emphasis in grade 6 on the number line, which is used to illustrate the relative relationships in quantities as well as positive and negative integers (NCDPI, 2013). Students may have leveraged their understanding of the number line when reasoning about the relative sizes of objects and may have been unable to recall the place value or measurement conversions from the previous grade, allowing for the differences in relationships across grades 5 and 6. The relationship between students' mathematics achievement and both relative and absolute scale strengthened with increasing

grade level in grades 7 and 8, indicating that as students get older, those students with a stronger sense of size and scale are likely to have higher mathematics achievement. This may be due to the emphasis on more abstract concepts in mathematics in late middle grades, indicating that students who can perform more abstract mathematics are also likely to have strong concepts of size and scale.

Reading Comprehension. Another explanation may be that there is an underlying reading ability that may explain the relationship between performance on scale assessments and science achievement, since both instruments rely on the students' ability to read at grade level to understand the assessment items. For example, students must be able to read and comprehend terms such as bacteria, virus, hallway, and galaxy on the scale assessments in order to identify both relative and absolute sizes. Likewise, the written format of the science EOG requires students can read and answer multiple choice questions related to science that require some knowledge of science vocabulary and reading comprehension. Though not examined in this study, it would be of benefit to include students' achievement on reading assessments in future work that investigates relationships such as these.

Developmental Underpinnings. Cognitive development may be another explanation for the trends observed in the relationship between students' concepts of scale and students' academic achievement. Though there is some debate as to the underlying processes that lead to adolescents' cognitive development, there is general consensus that during early adolescence individuals show notable improvements in reasoning, information processing, and expertise (Keating, 2004). As such, by grade 8, students may be more capable of abstract reasoning that may allow for a strengthened understanding of concepts of scale,

proportion, and quantity. This may help to explain the shift in the relationship between relative and absolute concepts of scale and academic achievement. In grade 6 students' understanding of relative size and scale were more significantly correlated with academic achievement than students' concepts of absolute size and scale. In contrast, in grades 7 and 8, the relationship shifted, with students' concepts of absolute scale being more strongly correlated to students' academic achievement. Over time, students' concepts of both relative and absolute size and scale improved (see appendix C), with performance (measured by percent accuracy) on the relative scale assessments consistently higher than performance on absolute scale assessments. Taken together, this may indicate that as science and mathematics content moves away from the human scale and into more abstract concepts, as noted in the science and mathematics standards, students with a stronger sense of absolute scale are more likely to have higher achievement in science and mathematics. This may have developmental underpinnings as these results support prior research in learning trajectories of size and scale (Delgado, 2009; Jones & Taylor, 2009; Tretter et al., 2006) in which individuals from different ages and levels of expertise acquired concepts of relative scale prior to concepts of absolute scale. It may be that because students in lower middle grades (i.e., grade 6) have a significantly stronger understanding of relative scale than absolute scale ($Z = 2.88, p < 0.05$) these individuals tend to leverage relative scale when reasoning about mathematics more so than absolute scale, though this did not hold for the relationship in grade 5/pre-6 results. In contrast, in grades 7 and 8, students may tend to rely on absolute scale when reasoning about mathematics and science as indicated by the greater strength of the relationship for absolute scale than relative scale.

The developmental underpinnings of understanding size and scale as well as the results of this study are supported by neuropsychological research that has investigated individuals' sense of numeracy. The ability to perceive approximate quantities is supported by a system in the brain called the Approximate Number Systems (ANS) (Libertus, 2015; Xu & Spelke, 2000). Humans and non-human animals share the innate ability to discriminate between approximate quantities quickly and without counting (i.e., which bush contains a greater number of berries) (Feigenson, Libertus, & Halberda, 2013). This ability to perceive relative quantities is developmental, as improvement of ANS acuity is most dramatic in young children, slows in middle school, and plateaus by age 30 (Halberda et al., 2008). The innate nature of the ANS may explain the presence of a stronger correlation between achievement and relative size and scale compared to absolute size and scale in younger grades because students are born with a sense of relative size (i.e., ANS acuity) and may only develop a sense of absolute scale following the acquisition of more formal mathematics. For example, a student may leverage her ANS to state that the length of a football field is less than the distance across North Carolina by discriminating between relative quantities; however, to assign absolute sizes to each of the aforementioned objects requires not only the ability to state that one is larger or smaller than the other, but also assign symbolic numerical quantities and units to each distance (i.e., 100 meters or 800 kilometers). Thus, students in younger grades may leverage their ANS to discriminate between relative quantities as demonstrated by the strong relationship between students' achievement and students' sense of relative scale at early ages. In contrast, as students develop a sense of absolute scale, they are more likely to rely on formal mathematics, acquired with each additional year of

schooling, which may support the development of a sense of absolute scale. Furthermore, the ANS may also support the ability to estimate quantity, as students who have a more acute ANS may be better at estimating the size of objects because they are better able to discriminate between quantities that are close in number. For example, a student that can discriminate between 14 and 16 objects (a ratio of 7:8) may be better at comparing the sizes of a bacterium and a red blood cell. While recent research has examined the relationship between science achievement and ANS acuity (see Delgado, Jones, You, Robertson, Chesnutt & Halberda, 2017) and found no significant correlation, the study sample size was relatively small ($N = 29$) and warrants a closer look.

Future Areas of Research. Based on the results of this study, there is clearly a positive and significant relationship between students' concepts of size and scale and students' achievement in mathematics, and this relationship holds across grades 5 through 8. Likewise, there is a strong, significant, positive correlation between students' concepts of size and scale and students' achievement in science in grades 5 and 8. However, the degree to which one factor explains the other is in need of further investigation. Because students' concepts of size and scale were positively correlated to academic achievement, this study provides an argument for a closer examination of contributing factors that may play a role in students' understanding of size and scale such as out of school science experiences, targeted instruction in size and scale, or general ability in mathematics or science. This research provides support for the three-dimensional pedagogy proposed by the NRC and NGSS, and offers support for both standards and pedagogical practices that could allow for deep learning in science and mathematics.

References

- Albarracín, L., & Gorgorió, N. (2014). Devising a plan to solve fermi problems involving large numbers. *Educational Studies in Mathematics*, 86(1), 79-96. doi:10.1007/s10649-013-9528-9
- American Association for the Advancement of Science. (1994). *Benchmarks for science literacy* Oxford University Press.
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 42(1), 189-201. doi:10.1037/0012-1649.41.6.189
- Bybee, R. W. (2014). NGSS and the next generation of science teachers. *Journal of science teacher education*, 25(2), 211-221.
- Bybee, R. (2015). Scientific literacy. In R. Gunstone (Ed.), *Encyclopedia of science education* (pp. 944-947) Springer Netherlands. Retrieved from http://link.springer.com/referenceworkentry/10.1007/978-94-007-2150-0_178
- Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, 133(1), 59-68. doi:10.1162/001152604772746701
- Cheek, K. A. (2013). How geoscience novices reason about temporal duration: The role of spatial thinking and large numbers. *Journal of Geoscience Education*, 61(3), 334-348. doi:10.5408/12-365.1

- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Coleman. (1981). *Public and private schools. an analysis of high school and beyond: A national longitudinal study for the 1980's*. Washington. D.C.: National Center for Educational Statistics.
- DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of research in science teaching*, 37(6), 582-601.
- Delgado, C. (2009). Development of a Research-Based Learning Progression for Middle School Through Undergraduate Students' Conceptual Understanding of Size and Scale (Doctoral dissertation), The University of Michigan, Ann Arbor, MI.
- Delgado, C. (2014). Collective landmarks for deep time: A new tool for evolution education. *Journal of Biological Education*, 48(3), 133-141. doi:10.1080/00219266.2013.849280
- Delgado, C., Jones, M. G., You, H. S., Robertson, L., Chesnutt, K., & Halberda, J. (2017). Scale and the Evolutionarily-Based Approximate Number System: An Exploratory Study. *International Journal of Science Education*, 1-17.
- Efthimiou, C. J., & Llewellyn, R. A. (2007). Cinema, fermi problems and general education. *Physics Education*, 42(3), 253. doi:10.1088/0031-9120/42/3/003
- Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives*, 7(2), 74-79. doi:10.1111/cdep.12019

- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational Research: An Introduction* (8th ed.). New York, NY: Pearson.
- Golledge, R. G., Gale, N., Pellegrino, J. W., & Doherty, S. (1992). Spatial knowledge acquisition by children: Route learning and relational distances. *Annals of the Association of American Geographers*, 82(2), 223-244. doi:10.1111/j.1467-8306.1992.tb01906.x
- Handley, H. M., & Morse, L. W. (1984). Two-year study relating adolescents' self-concept and gender role perceptions to achievement and attitudes toward science. *Journal of Research in Science Teaching*, 21(6), 599-607. doi:10.1002/tea.3660210606
- Hegarty, M. (2014). Spatial thinking in undergraduate science education. *Spatial Cognition & Computation*, 14(2), 142-167. doi:10.1080/13875868.2014.889696
- Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, 34(2), 151-176. doi:10.1016/j.intell.2005.09.005
- Hogan, T. P., & Brezinski, K. L. (2003). Quantitative estimation: One, two, or three abilities? *Mathematical Thinking and Learning*, 5(4), 259-280. doi:10.1207/S15327833MTL0504_02
- Jones, M. G., Forrester, J. H., Robertson, L. E., Gardner, G. E., & Taylor, A. R. (2012). Accuracy of estimations of measurements by students with visual impairments. *Journal of Visual Impairment & Blindness*, 106(6), 351-355.

- Jones, M. G., Gardner, G., Taylor, A. R., Wiebe, E., & Forrester, J. (2011). Conceptualizing magnification and scale: The roles of spatial visualization and logical thinking. *Research in Science Education, 41*(3), 357-368. doi:10.1007/s11165-010-9169-2
- Jones, M. G., & Rua, M. J. (2008). Conceptual representations of flu and microbial illness held by students, teachers, and medical professionals. *School Science and Mathematics, 108*(6), 263-278. doi:10.1111/j.1949-8594.2008.tb17836.x
- Jones, M. G., & Taylor, A. R. (2009). Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching, 46*(4), 460-475. doi:10.1002/tea.20288
- Jones, G., Taylor, A., & Broadwell, B. (2009). Estimating linear size and scale: Body rulers. *International Journal of Science Education, 31*(11), 1495-1509. doi:10.1080/09500690802101976
- 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. doi:10.1007/s10956-006-9034-2
- 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. doi:10.1080/09500690701416624
- Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. *Review of Educational Research, 68*(4), 413-449. doi:10.2307/1170734

- Krajcik, J., Codere, S., Dahsah, C., Bayer, R., & Mun, K. (2014). Planning instruction to meet the intent of the Next Generation Science Standards. *Journal of Science Teacher Education*, 25(2), 157-175.
- Liben, L. S., Kastens, K. A., & Christensen, A. E. (2011). Spatial foundations of science education: The illustrative case of instruction on introductory geological concepts. *Cognition and Instruction*, 29(1), 45-87. doi:10.1080/07370008.2010.533596
- Liben, L. S., Myers, L. J., Christensen, A. E., & Bower, C. A. (2013). Environmental-scale map use in middle childhood: Links to spatial skills, strategies, and gender. *Child Development*, 84(6), 2047-2063. doi:10.1111/cdev.12090
- Liu, O. L., Ryoo, K., Linn, M. C., Sato, E., & Svihla, V. (2015). Measuring knowledge integration learning of energy topics: A two-year longitudinal study. *International Journal of Science Education*, 37(7), 1044-1066. doi:10.1080/09500693.2015.1016470
- Magana, A. J., Brophy, S. P., & Bryan, L. A. (2012). An integrated knowledge framework to characterize and scaffold size and scale cognition (FS2C). *International Journal of Science Education*, 34(14), 2181-2203. doi:10.1080/09500693.2012.715316
- National Research Council. (1996). *National science education standards*. [S.l.]: National Academy Press.
- National Research Council. (2012). *A framework for K-12 science education electronic resource: Practices, crosscutting concepts, and core ideas*. Washington, D.C.: The National Academies Press.

North Carolina Department of Public Instruction (n.d.-a). *North Carolina Essential Standards 3-5 Science*. Retrieved from <http://scnces.ncdpi.wikispaces.net/file/view/NCES+3-5.pdf/565466269/NCES%203-5.pdf>

North Carolina Department of Public Instruction (n.d.-b). *North Carolina Essential Standards 6-8 Science*. Retrieved from http://scnces.ncdpi.wikispaces.net/file/view/NCES_Science6-8.pdf/565151883/NCES_Science6-8.pdf

North Carolina Department of Public Instruction. (2013). *6th Grade Mathematics Unpacked Content*. Retrieved from <http://www.dpi.state.nc.us/docs/curriculum/mathematics/scos/6.pdf>

North Carolina Department of Public Instruction. (2014). *Interpretive Guide to the WinScan Score Reports for the North Carolina End-of-Grade Assessments*. Retrieved from <http://www.ncpublicschools.org/docs/accountability/testing/technotes/1415eogwsguide.pdf>

North Carolina Department of Public Instruction. (2015). *5th Grade Mathematics Unpacked Content*. Retrieved from <http://www.ncpublicschools.org/docs/curriculum/mathematics/scos/5.pdf>

North Carolina Report Card. (2015). Retrieved from <https://ncreportcards.ondemand.sas.com/landing.html>

NGSS Lead States. (2013). *Next generation science standards: For states, by states*.

- Washington, US: National Academies Press. Retrieved from
<http://site.ebrary.com/lib/ncsu/docDetail.action?docID=10863742&ppg=1>
- Piaget, J. & Inhelder, B. (2013). *Mental Imagery in the Child: Selected Works* (Vol. 6) New York, NY: Routledge
- Piaget, J., Inhelder, B., & Szeminska, A. (Eds.). (1960). *The child's conception of geometry*. New York: Basic Books. Retrieved from
http://buprimo.hosted.exlibrisgroup.com/primo_library/libweb/action/dlSearch.do?institution=BOSU&vid=BU&query=any,contains,99195692400001161
- Preacher, K. J. (2002, May). Calculation for the test of the difference between two independent correlation coefficients [Computer software]. Available from <http://quantpsy.org>.
- Resnick, I., Davatzes, A., Newcombe, N. S., & Shipley, T. F. (2016). Using relational reasoning to learn about scientific phenomena at unfamiliar scales. *Educational Psychology Review*, 29(1), 11-25. doi:10.1007/s10648-016-9371-5
- Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93(3), 604-614. doi:10.1037/0022-0663.93.3.604
- Simpson, R. D., & Steve Oliver, J. (1990). A summary of major influences on attitude toward and achievement in science among adolescent students. *Science Education*, 74(1), 1-18. doi:10.1002/sce.3730740102

- Stage, E. K., Asturias, H., Cheuk, T., Daro, P. A., & Hampton, S. B. (2013). Opportunities and challenges in next generation standards. *Science*, *340*(6130), 276-277.
- Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. *Philosophy of science, cognitive psychology, and educational theory and practice*, 176.
- Taylor, A., & Jones, G. (2009). Proportional reasoning ability and concepts of scale: Surface area to volume relationships in science. *International Journal of Science Education*, *31*(9), 1231-1247. doi:10.1080/09500690802017545
- 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*(2), 191-221. doi:AID-TEA1003>3.0.CO;2-C
- Tretter, T. R., Jones, M. G., Andre, T., Negishi, A., & Minogue, J. (2006). Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching*, *43*(3), 282-319. doi:10.1002/tea.20123
- Tretter, T. R., Jones, M. G., & Minogue, J. (2006). Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching*, *43*(10), 1061-1085. doi:10.1002/tea.20155
- Vandelacluze, V. (2012). The world in their minds: A multi-scale approach of children's representations of geographical space. In L. Zentai, & J. R. Nunez (Eds.), *Maps for the future* (pp. 229-241) Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-19522-8_19

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817-835. doi:10.1037/a0016127


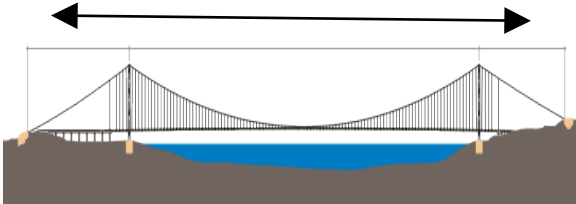
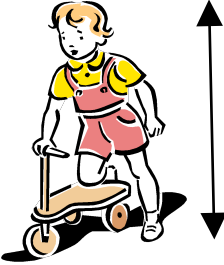


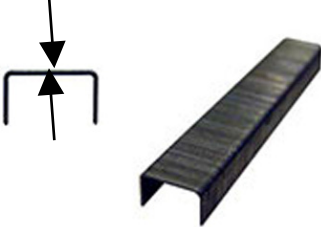
Appendices

Appendix A: Scale Card Sort

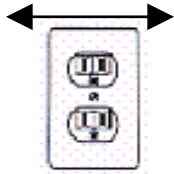
Scale Card Sort (SCS) instructions

Please sort these cards into piles according to the size depicted on each card. Objects and distances similar in size to each other should go into the same pile. You can make as many piles as you like and have as many cards in each pile as you like – you don't need the same number of cards in each pile.

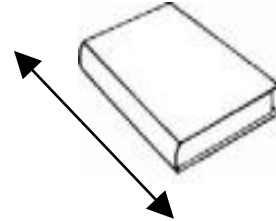
When you have completed your piles, arrange the cards from smallest size depicted on the card to largest size depicted on the card.

<p>Length of an SUV (Sport Utility Vehicle)</p> 	<p>Span of the longest bridge in the world from post to post</p> 
<p>Height of a typical 5 year old child</p> 	<p>Height of a typical NBA basketball player</p> 
<p>Diameter of a quarter</p> 	<p>Thickness of a staple</p> 

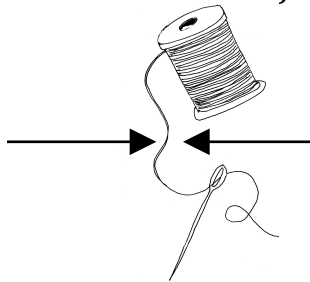
Width of a electrical outlet cover



Length of a typical phone book



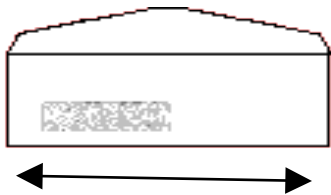
Thickness of sewing thread



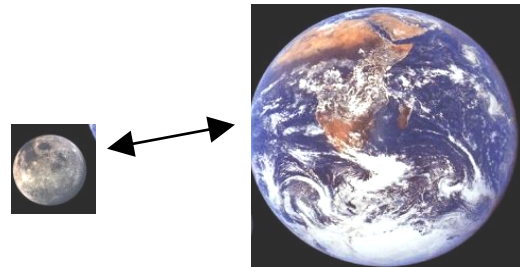
Diameter of the Earth



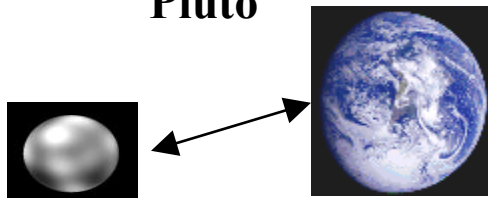
Length of a business envelope



Distance from Earth to the Moon



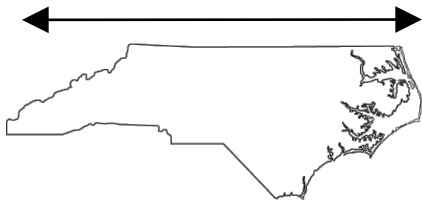
Distance from Earth to Pluto



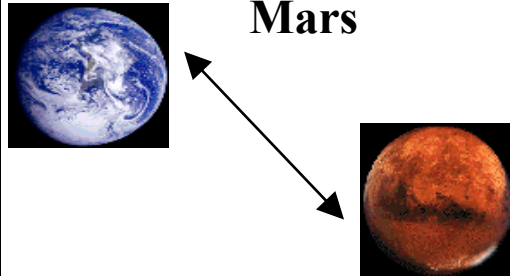
Distance from Miami to Boston



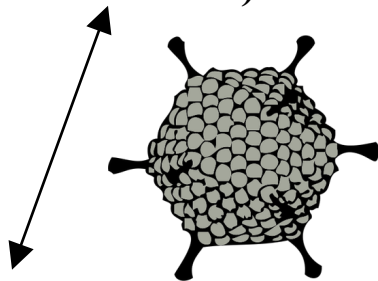
East –to –West length of North Carolina



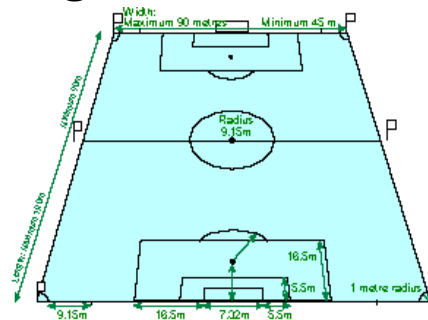
Distance from Earth to Mars



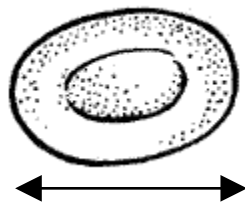
Virus (adenovirus shown below)



Length of a soccer field



Diameter of a red blood cell



Water Molecule



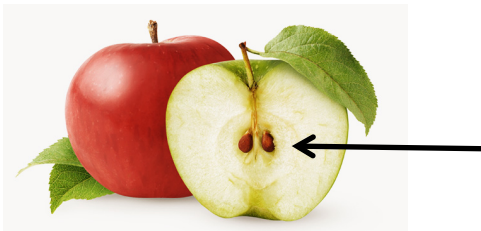
Width of a typical wedding ring



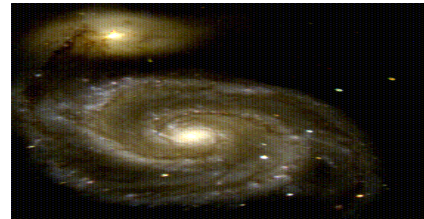
Length of a main hallway in a typical school



Length of an apple seed



Distance from the Milky Way galaxy to the nearest galaxy

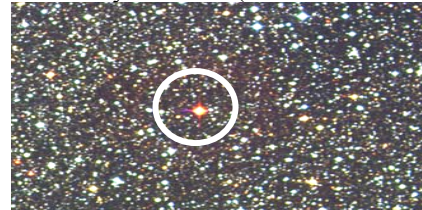


Length of a typical bedroom



Distance from sun to the nearest star

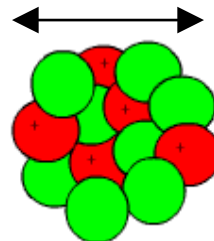
Nearest star in the sky is circled (Proxima Centauri)



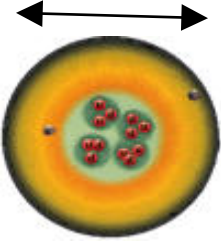
Altitude of the Space Station above the Earth



Nucleus of an oxygen atom




Diameter of a helium atom



A diagram of a helium atom showing a central nucleus with two red protons and two blue neutrons, surrounded by two concentric electron shells. A horizontal double-headed arrow is positioned above the atom to indicate its diameter.

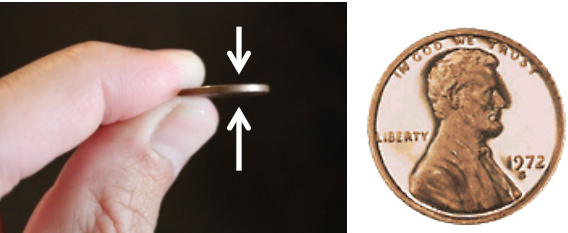
Distance from
the Milky Way galaxy to the *farthest* known galaxy

Each spot of light is not one star, but rather a whole galaxy of stars as seen through the Hubble Telescope



A photograph showing a vast field of distant galaxies, appearing as numerous bright, multi-colored spots of light against a dark background, representing the farthest known galaxies from the Milky Way.

Thickness of a penny
(How much it rises above the card's surface is the thickness)



A photograph of a hand holding a penny between the thumb and index finger. Two white arrows point vertically towards each other, indicating the thickness of the coin. To the right is a separate image of a penny showing the profile of Abraham Lincoln, with the words 'IN GOD WE TRUST', 'LIBERTY', and the year '1972' visible.

Student ID

Appendix B: Comparison of Student Mean Scores Between Students that Completed Less than Half of All Assessments and Students that Completed at Least Half of All Assessments

Table B1.

Descriptive statistics for EOG assessments between groups of students

Assessment	Student		Mean	Std. Deviation	Std. Error
	Groups	N			
Grade 5 Science EOG	A	60	261.88	31.12	4.02
	B	155	257.07	23.94	1.92
Grade 5 Mathematics EOG	A	60	430.02	64.26	8.30
	B	156	444.70	39.72	3.18
Grade 6 Mathematics EOG	A	62	437.71	51.49	6.54
	B	165	440.25	44.34	3.45
Grade 7 Mathematics EOG	A	63	449.29	11.35	1.43
	B	165	449.93	10.05	0.78
Grade 8 Mathematics EOG	A	66	448.85	10.77	1.33
	B	165	448.92	10.08	0.78
Grade 8 Science EOG	A	66	251.06	10.76	1.33
	B	165	251.94	9.20	0.72

Note. A = Students who completed less than half of all assessments, B = Students who completed at least half of all assessments, 14 total assessments

Table B2.

Descriptive statistics for scale assessments between groups of students who completed less than half or at least half of all 14 assessments

Assessment	Student Groups	N	Mean	Std. Deviation	Std. Error Mean
Grade 6 SCS Pre	A	40	15.86	9.40	1.49
	B	113	16.30	9.08	0.85
Grade 6 SOQ Pre	A	67	11.69	5.50	0.67
	B	191	11.35	5.69	0.41
Grade 6 SCS Post	A	41	18.43	9.49	1.48
	B	114	16.99	9.30	0.87
Grade 6 SOQ Post	A	45	11.93	6.20	0.92
	B	127	13.76	5.63	0.50
Grade 7 SCS Post	A	36	24.00	6.38	1.06
	B	100	21.36	7.45	0.74
Grade 7 SOQ Post	A	36	15.64	5.26	0.88
	B	102	14.72	5.59	0.55
Grade 8 SCS Post	A	66	18.19	8.42	1.04
	B	164	18.95	8.34	0.65
Grade 8 SOQ Post	A	66	14.22	6.08	0.75
	B	172	14.28	6.19	0.47

Note. SCS = Scale Card Sort, SOQ = Scale of Objects Questionnaire, A = Students who completed less than half of all assessments, B = Students who completed at least half of all assessments

Table B3

Independent T-test of students' EOG scores between groups of students who completed less than half or at least half of all 14 assessments

	Equal Variance	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Grade 5 Science EOG	A	2.79	0.10	1.21	213.00	0.23	4.81	3.97	-3.02	12.64
	B			1.08	87.36	0.28	4.81	4.45	-4.04	13.67
Grade 5 Mathematics EOG	A	14.89	0.00	-2.02	214.00	0.04	-14.69	7.26	-28.99	-0.39
	B			-1.65	76.98	0.10	-14.69	8.88	-32.38	3.00
Grade 6 Mathematics EOG	A	0.85	0.36	-0.37	225.00	0.71	-2.54	6.91	-16.16	11.07
	B			-0.34	96.92	0.73	-2.54	7.39	-17.22	12.13
Grade 7 Mathematics EOG	A	2.76	0.10	-0.42	226.00	0.68	-0.65	1.54	-3.69	2.39
	B			-0.40	101.22	0.69	-0.65	1.63	-3.88	2.59
Grade 8 Mathematics EOG	A	0.31	0.58	-0.05	229.00	0.97	-0.07	1.50	-3.02	2.88
	B			-0.04	113.04	0.97	-0.07	1.54	-3.12	2.99
Grade 8 Science EOG	A	3.05	0.08	-0.62	229.00	0.53	-0.88	1.41	-3.65	1.90
	B			-0.58	104.99	0.56	-0.88	1.51	-3.87	2.11

Note. A = Equal variance assumed, B = Equal variances not assumed

Table B4

Independent T-test of student's scale assessment scores between groups of students who completed less than half or at least half of all 14 assessments

	Equal Variance	F	Sig.	t	df	Sig. (2- tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									<u>Lower</u>	<u>Upper</u>
Grade 6 SCS Pre	A	0.37	0.55	-0.26	151.00	0.80	-0.44	1.69	-3.77	2.89
	B			-0.26	66.46	0.80	-0.44	1.71	-3.86	2.98
Grade 6 SOQ Pre	A	0.48	0.49	0.43	256.00	0.67	0.34	0.80	-1.24	1.92
	B			0.43	119.11	0.67	0.34	0.79	-1.22	1.90
Grade 6 SCS Post	A	0.04	0.84	0.85	153.00	0.40	1.44	1.70	-1.92	4.80
	B			0.84	69.49	0.41	1.44	1.72	-1.99	4.87
Grade 6 SOQ Post	A	3.43	0.07	-1.83	170.00	0.07	-1.83	1.00	-3.81	0.15
	B			-1.74	71.37	0.09	-1.83	1.05	-3.92	0.26

Note. A = Equal variance assumed, B = Equal variances not assumed

Table B4 Continued

	Equal Variance	F	Sig.	t	df	Sig. (2- tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									<u>Lower</u>	<u>Upper</u>
Grade 7 SCS Post	A	1.88	0.17	1.89	134.00	0.06	2.65	1.40	-0.12	5.41
	B			2.04	71.72	0.05	2.65	1.30	0.06	5.23
Grade 7 SOQ Post	A	0.53	0.47	0.86	136.00	0.39	0.92	1.07	-1.19	3.04
	B			0.89	64.90	0.38	0.92	1.04	-1.15	2.99
Grade 8 SCS Post	A	0.03	0.85	-0.62	228.00	0.53	-0.76	1.22	-3.16	1.64
	B			-0.62	119.09	0.54	-0.76	1.22	-3.18	1.66
Grade 8 SOQ Post	A	0.07	0.79	-0.07	236.00	0.94	-0.07	0.89	-1.82	1.69
	B			-0.07	119.62	0.94	-0.07	0.89	-1.82	1.69

Note. A = Equal variance assumed, B = Equal variances not assumed

**Appendix C: Descriptive statistics for all assessments in all years of the study for
all students**

Table C1.

Descriptive statistics for performance on all assessments in all years.

	N	Minimum	Maximum	Mean	Std. Deviation
Grade 5 Science EOG	215	224	366	258.414	26.15573
Grade 5 Mathematics EOG	216	240	475	440.6248	48.10621
Grade 6 Mathematics EOG	227	240	473	439.5595	46.29685
Grade 7 Mathematics EOG	228	431	473	449.7544	10.40125
Grade 8 Mathematics EOG	231	431	474	448.8961	10.25787
Grade 8 Science EOG	231	228	273	251.6883	9.65752
Grade 6 SCS Pre	153	0	31	16.1863	9.13205
Grade 6 SOQ Pre	258	0.5	22	11.4341	5.63338
Grade 6 SCS Post	155	0	30	17.3677	9.34148
Grade 6 SOQ Post	172	1	23	13.2849	5.82018
Grade 7 SCS Post	136	1	31	22.0551	7.25276
Grade 7 SOQ Post	138	1	24.5	14.9565	5.50364
Grade 8 SCS Post	230	0	31	18.7304	8.35113
Grade 8 SOQ Post	238	1	23.5	14.2668	6.14498

Students' Concepts of Size and Scale: Factors of Influence

Abstract

This study examined the degree to which individual differences in students' (N=232) concepts of size and scale are explained by factors such as students' innate sense of number, out-of-school science experiences, exposure to size and scale instruction, gender identities, and racial/ethnic identities. Findings indicated students' experiences with scale outside of school, exposure to size and scale instruction, and racial/ethnic identities statistically and significantly added to the prediction model. Results from this study can both inform the movement towards incorporating crosscutting concepts into pedagogy as well as inform educators, administrators, and other stakeholders of the factors that may shape students' understanding of the cross cutting concept of scale, proportion, and quantity.

Students' Concepts of Size and Scale: Factors of Influence

National K-12 science standards have encouraged teaching using broad themes as a means of uniting science content, integrating other STEM areas such as mathematics and engineering, and providing students with deep and meaningful science learning experiences (Next Generation Science Standards [NGSS] Lead States, 2013). The crosscutting concept of *scale, proportion and quantity*, often called *size* and *scale* in the literature, is one such theme (National Research Council [NRC], 2012; NGSS Lead States, 2013). Magana, Brophy, and Bryan (2012) defined *size* as bulk or quantity, and Resnick, Davatzes, Newcombe, and Shipley (2016) defined *scale* as the systems that provide a basis by which relative sizes can be compared. New research has suggested that a significant, positive relationship exists between students' concepts of size and scale and students' achievement in both science and mathematics during middle school (Chesnutt, Jones, Corin, Hite, Childers, Cayton, & Ennes, in preparation). The correlational findings in the study by Chesnutt et al. warrants a closer look at specific factors that may predict students' concepts of size and scale. This investigation examined the proportion of the variation in students' performance on size and scale assessments that could be explained by five factors: science scale capital (described in the sections that follow), students' exposure to size and scale instruction, students' innate sense of number, students' gender identities, and students' racial/ethnic identities. This investigation examined eighth grade students' (N=232) concepts of size and scale and sought to address the following research questions:

To what degree do the following factors explain the variation in students' concepts of size and scale:

- a) students' science scale capital*
- b) students' exposure to size and scale instruction*
- c) students' innate sense of approximate number*
- d) students' gender identities, and*
- e) students' racial/ethnic identities*

Prior research on individuals' concepts of size and scale focused primarily on reporting individuals' understandings of size and scale (e.g., Tretter et al., 2006) or was limited to correlational analyses (e.g., Chesnutt et al., in preparation; Delgado, Jones, You, Robertson, Chesnutt, & Halberda, 2017). This research builds on previous studies in that it seeks to examine factors that may explain the variance observed in students' concepts of size and scale. Understanding the degree to which these factors can predict students' concepts of size and scale can help to inform educators, policy makers, and other stakeholders of instructional resources that may foster deep learning and higher achievement in science and mathematics.

Literature Review

Crosscutting Concepts. The National Research Council identified disciplinary core ideas, science and engineering practices, and crosscutting concepts as components of a three-dimensional pedagogical approach aimed at providing students with opportunities for deep learning in science (NRC, 2012). As part of this pedagogy, each of the three components contributes to instructional strategies that encourage students to meet performance

expectations outlined in the NGSS (NGSS Lead States, 2013). Different from discrete facts found in previous science standards, the performance expectations outlined in the NGSS require students to integrate disciplinary core ideas, crosscutting concepts, and science and engineering practices as they construct understanding of scientific phenomena (NGSS Lead States, 2013). Students' understanding of crosscutting concepts such as size and scale have demonstrated a positive, significant correlation with students' science and mathematics achievement (Chesnutt et al., in preparation), providing some evidence to the claims that these broad themes may be fundamental to learning in science and supporting proposals for a closer examination of crosscutting concepts in science education. Prior research has examined individual differences in concepts of size and scale across age and expertise (Delgado, 2009; Jones & Taylor, 2009; Tretter, Jones, Andre, et al., 2006) as well as the relationship between individuals' understanding of size and scale and individuals' exposure to size and measurement instruction (Jones, Taylor, & Broadwell, 2009), experiences in professional settings (Jones & Taylor, 2009), and science or mathematical abilities (Delgado et al., 2017; Chesnutt et al., in preparation). Absent from the literature is an examination of the degree to which multiple factors (e.g., instruction, science scale capital, academic performance, gender, and race/ethnicity) considered in a single model can predict students' concepts of size and scale. The following review of the literature summarizes what is known about individuals' concepts of size and scale as it relates to science education and describes research on out of school experiences related to scale, intervention studies in size and scale, and research that examined factors of gender or race as it relates to understanding size and scale.

Individual Differences in Concepts of Size and Scale. Individuals' understanding of size and scale is often determined through measurement estimation tasks (e.g., Tretter et al., 2006; Delgado et al., 2017; Jones & Taylor, 2009) because these measurement estimations provide a deep and intuitive understanding of an individual's perceptions of the relationship between physical space and representative numerical symbol (Albarracín & Gorgorió, 2014). Essential to understanding a cross cutting science theme such as size and scale, are the mathematical concepts of quantity and proportion, as quantity allows for representation of magnitude (Magana et al., 2012) and proportional reasoning allows for the conceptualization of the equivalence of appropriate scalar ratios (Lamon, 1993). Children learn about quantity, in part, by mapping a number of physical objects to a quantitative symbol (Booth & Siegler, 2006; Gallistel & Gelman, 1992; Joram, Subrahmanyam, & Gelman, 1998; Le Core & Carey, 2007). Measurement tasks have been used to index students' concepts of size and scale because they allow for the mapping of physical objects to a symbolic number, indexing not only the ability to estimate the magnitude of an object but also the ability to compare magnitudes using scales (Jones et al., 2009).

Early research on size and scale included investigations of the conceptualization of mapping and navigation using linear scales (e.g., Golledge, Gale, Pellegrino, & Doherty, 1992). Navigation and mapping research later shifted from linear scales to a focus on spatial ability and spatial cognition (e.g., Hegarty, 2014; Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Liben, Kastens, & Christensen, 2011). More recently, science education researchers have begun to examine the crosscutting theme of size and scale on students' concepts of geologic time. For example, Trend (2001) found that students categorized time

into extremely ancient, moderately ancient, or less ancient. Furthermore, students were more familiar with events or landmarks at the extremes of geologic time such as the formation of the Earth or more recent human history (Trend, 2001; Delgado, 2014), and students were most familiar with events that occurred during human history (Delgado, 2014). More recently, the examination of students' perceptions of the relationship between spatial distance and geologic time duration revealed that undergraduate students (N=17) were apt to confuse large magnitudes (e.g., million, billion), misrepresent temporal and spatial relationships (e.g., short time periods were represented as too long spatially), and not use events or landmarks when reasoning about time (Cheek, 2013). The misrepresentation of time periods was supported by other researchers who found that students often overestimated the size of small objects and underestimated the size of large objects (Delgado, 2009; Tretter, Jones, & Minogue, 2006). Similar to research on measurement estimation, which found that individuals are in general very poor at understanding size and scale (Joram et al., 1998), education researchers found that K-12 and undergraduate students had poor conceptualizations of both the sequence and duration of geologic time, suggesting calls to examine students' concepts of size and scale in other science contexts (Delgado, 2014; Trend, 2001).

Building on research pertaining to students' concepts of geologic time, was a movement to examine students' concepts of scale through measurement and measurement estimation. One of the earliest research studies in size and scale involving measurement estimation examined differences in understanding of size and scale for students in grades 5, 7, 9, 12, and doctoral students (Tretter, Jones, Andre et al., 2006). Tretter and colleagues

found that all participants (N=215) could more accurately estimate the relative size of objects than the actual size of objects and that this ability improved with age and expertise, which the researchers attributed to increased personal experiences with a variety of sizes and scales for older and more expert participants (Tretter, Jones, Andre et al., 2006). In conjunction with this study, Tretter, Jones, and Minogue (2006) studied differences in the same participants' concepts of size and scale between small and large extremes and found individuals were more accurate identifying the actual and relative sizes of larger extremes (e.g., solar system) than smaller extremes (e.g., virus) with the greatest accuracy observed at the human scale. These findings are supported by research in geologic time scales, which found that individuals' concepts of geologic time decreased the farther an event occurred from human history (Delgado, 2014; Trend, 2001) as well as later research in teachers' (Jones, Tretter, Taylor, & Oppewal, 2008), professionals' (Jones & Taylor, 2009) and students' (Delgado, 2009; Jones, Taylor, et al., 2007; Jones et al., 2009; Resnick et al., 2016) concepts of size and scale. A series of research studies followed these initial investigations of concepts of size and scale, which revealed several trends in how individuals acquire, reason about, and make use of a sense of size and scale (Cheek, 2013; Jones et al., 2008; Jones & Taylor, 2009; Jones, Taylor, & Broadwell, 2009; Resnick et al., 2016).

Because individuals generally have a poor sense of size and scale (outside of the human-scale), it was of interest to researchers to examine the ways in which those individuals with a strong sense of scale acquired this understanding. Tretter, Jones, Andre, et al. (2006) reported that personal experiences with specific sizes and scales likely contributed to experts' understanding of size and scale. As revealed through interviews, Jones and Taylor

(2009) found that kinesthetic experiences, specifically, likely contributed to professionals' acquisition of a sense of size and scale.

Another approach to researching size and scale has been to examine how individuals make meaning of different sizes or reason about different scales. When reasoning about size and scale, most all studies found that the human sizes and scales were the most acute and were often used as anchors or landmarks for size comparisons (Cheek, 2013; Jones, Taylor, et al., 2007; Jones et al., 2008; Jones, Taylor, & Broadwell, 2009; Resnick et al., 2016; Tretter, Jones, Andre, et al., 2006; Tretter, Jones, & Minogue, 2006). In addition, researchers found that the number of anchors or landmarks at different scales that participants used during their reasoning increased with an improved sense of size and scale (Jones & Taylor, 2009; Tretter, Jones, Andre, et al., 2006). This was also supported by research in conceptualization of geologic time, which found that the number of events or landmarks (e.g., the separation of the supercontinent Pangaea) increased as individuals demonstrated a more acute sense of geologic time (Trend, 2001; Delgado, 2014). As individuals reason about different sizes and scales, they explain the process as mentally maneuvering using “jumps” (Tretter, Jones, & Minogue, 2006) and often times this can involve mathematical thought (Jones & Taylor, 2009).

A third approach to research on size and scale has been to examine how this conceptualization is used in professional settings. Jones and Taylor (2009) examined the relevance of size and scale in a variety of professions and found that individuals who used scale or their work (e.g., microbiologist, astronomer) reported that their jobs would be nearly impossible without a sound understanding of size and scale. Jones and Taylor (2009) also

found that the use of scale in the professional workplace was not limited to academic settings, as a range of professions required understanding of size and scale within specific domains (e.g., painters, builders). The ways in which individuals reason, acquire, and use a sense of size and scale can rely on many factors. While these aforementioned factors such as personal experiences or mathematical thought have been highlighted in research (e.g., Jones & Taylor, 2009; Jones, Taylor, et al., 2007), there is a need to understand the degree to which some factors reported as influential in understanding scale may interact together to play a role in the conceptualization of size and scale. These factors cannot be addressed without first examining the influence of culture on concepts of size and scale.

Measurement and estimation: a cultural phenomenon. The units or scales that individuals leverage to make measurements and measurement estimations are culturally derived (Jones & Taylor, 2009). The cultural basis of “unitizing” or the construction of contextualized units, has led researchers to conclude that cultural differences are the basis for differences in the accuracy with which individuals from varying cultures estimate measurements, and thus, understand size and scale (Delgado, 2013; Jones et al., 2010). For example, in an international comparison of teachers’ concepts of spatial scale between preservice and in-service teachers in the United States (U.S.), Austria, and Taiwan indicated that across all groups of teachers, concepts of scale increased with teaching expertise, as preservice teachers held less accurate concepts of scale than in-service teachers (Jones et al., 2010). Furthermore, U.S. teachers had less accurate concepts of scale than Austrian and Taiwanese teachers (Jones et al., 2010). Jones and colleagues (2010) proposed that the discrepancy across different groups of teachers might have been the result of cultural

differences between the types of measurement units adopted by each country. For example, Austria and Taiwan both use the metric system, providing individuals more personal experiences with scale in their daily lives, thus leading to a more robust understanding of scale (Jones et al., 2010). Delgado (2013) further investigated this idea by identifying the cross-cultural differences in students' concepts of scale as it related to the metric system. Delgado (2013) found that *Système International* (SI) natives (i.e., students for whom the metric system was part of their everyday lives) performed significantly better than SI-nonnatives (i.e., students in the U.S. who use U.S. Customary measurement system in everyday life) in estimating metric distances. The process of "unitizing" also improves with expertise, as experts are more likely to construct new units out of a need for relevant and convenient indices of measurement (Jones et al., 2006). For example, Grace Hopper, a pioneer in early computer science, is known for passing around a 12 inch piece of wire during her speaking engagements to represent a nanosecond, leading to the commonly used unit that electricity travels one foot per nanosecond (Wong & Stoyanov, 2006). Exercising "unitizing" as a means of contextualizing and simplifying complex sizes and scales is fundamental to individuals' concepts of spatial scales (Jones et al., 2006). The processes individuals use to mentally navigate between orders of magnitude such as unitizing (Jones et al., 2006), the use of landmarks (Trend, 2001; Delgado, 2014), and mental "jumps" (Tretter et al., 2006) has been documented in the science education research. However, leveraging only cultural differences fails to acknowledge other factors such as variation in access to resources or differences in quantitative ability, factors that can be constant across cultures.

There is a need for a closer examination of the degree to which individual differences in both cultural experiences (e.g., social interactions, resources, instruction) and non-cultural experiences (e.g., number sense) can predict students' conceptual understanding of size and scale. While prior research has examined the relationship between reasoning skills such as spatial visualization (Jones, Gardner, Taylor, Wiebe, & Forrester, 2011) and logical thinking (Jones et al., 2011; Taylor & Jones, 2009) and students' concepts of size and scale, this study includes factors not previously examined as potential explanatory variables in predicting students' concepts of size and scale. The following sections outline perspectives from which this study sought to examine concepts of size and scale; the first is science scale capital which includes a focus on out-of-school experiences and resources related to scale; the second is in-school instruction related to size and scale; the third is innate number sense which is an evolutionarily inherited ability supported by a system in the brain called the Approximate Number System; and the last section examines gender and racial/ethnic identities in science as it relates to concepts of size and scale.

Science Scale Capital. In order to support individuals' understanding of measurement, and thus concepts of size and scale, it is important to leverage cultural perspectives, viewpoints that are largely shaped by personal experiences (Delgado, 2013; Tretter, Jones, & Minogue, 2006; Jones & Taylor, 2009). The role of personal experiences in individuals' concepts of size and scale has been well documented in science education research (Jones et al., 2008; Jones & Rua, 2008; Jones & Taylor, 2009; Tretter, Jones, Andre, et al., 2006; Tretter, Jones, & Minogue, 2006). While efforts have been made to document the experiences that led to experts' concepts of size and scale, there is a need for a

quantitative investigation of the degree to which these experiences can predict concepts of size and scale understanding.

One specific category of personal experiences includes students' exposure to resources, events, or individuals that could enhance or support learning in science, particularly those experiences that take place when students are not in school. Out of school experiences provide a substantial amount of science learning for individuals of all ages (Falk & Dierking, 2010). In a study of professionals that use scale in their work, Jones and Taylor (2009) found that nearly all of the participants in their study (N=50) described out-of-school experiences in their youth as contributing to the development of their sense of scale.

Building on Bourdieu's idea of social capital (e.g., 1977; 1984) is the concept of *science capital* or the extent to which individuals have access to experiences, resources, and tools that promote science learning (Archer et al., 2012). One component of science capital is derived from family support. Archer and colleagues referred to *science family habits* as the degree to which families value and support students' science identities and science career aspirations (Archer et al., 2012). Because concepts of size and scale are based largely on personal experiences (Jones, Taylor, et al., 2007; Tretter et al., 2006), experiences which may not always take place in K-12 environments, there is a need for understanding the role *science capital* and *science habitus*, may play in students' understanding of crosscutting concepts such as scale, proportion, and quantity. Archer et al. (2012) found that students' *science capital* varied across different groups of students (with regard to race and class). The authors found that this difference in capital was related to differences in the sustained promotion and development of interest and aspirations in science which the authors attributed

to these families making science both paramount and frequent in their daily lives. Building on these findings, this research project proposes a new concept, *science scale capital*, which refers to students' access to experiences and resources that promote students' understanding of size and scale, and posits that differences in students' *science scale capital* may provide a possible source of the variation that exists across students' concepts of size and scale. Access to science scale capital may include *experiences* such as building, sewing, timing a race, estimating the distance or time for sporting events, travel, or attending science summer camps. These experiences may be enhanced to support a stronger science scale capital through *resources* such as adults with STEM careers, a family support system that promotes science, or scientific measurement and observational tools that are available outside of school such as meter sticks, thermometers, or binoculars. The model in Figure 1 illustrates that it is through the successful interaction of *experiences*, *resources*, and *tools* that students are able to gain access to science scale capital.

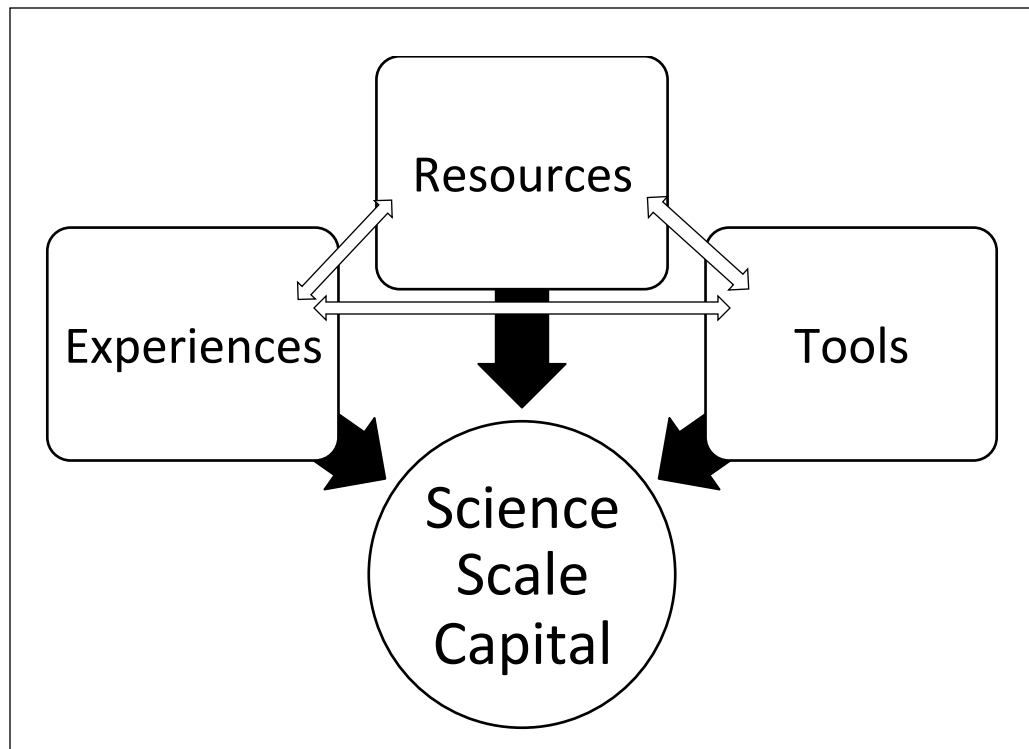


Figure 1. The model proposed for how individuals might gain access to science scale capital. It is through the interaction of *experiences* (such as building a model to scale), *resources* (such as adults with STEM experiences), and measurement *tools* (such as a thermometer, ruler, etc.) that science scale capital is acquired.

Size and Scale Intervention Studies. Efforts have been made to evaluate the efficacy of instruction or targeted interventions related to size and scale and the benefits of that instruction for both students and teachers. Jones and colleagues (Jones, Taylor, et al., 2007) found showing the film *Powers of Ten* to middle school students (n=22) significantly improved accuracy on card sorting tasks pre to post. Though not significant, students' abilities to correctly pair metric sizes in scientific notation to the corresponding metric scale also improved pre to post. Teachers in the study (n=6) reported that the film was a useful pedagogical tool, but researchers lamented that their data could not explain *which* aspects of

the film (e.g., zooming, focusing on the emptiness of space) were the keys to the film's effectiveness (Jones, Taylor, et al., 2007). Furthermore, the relatively small sample size (N=28) and all female sample population raise questions about the generalizability of the study to larger, more representative populations.

Building on research which found personal experiences, familiarity with the human scale, and kinesthetic experiences as factors which had been reported to help develop a sense of scale (Jones & Taylor, 2009; Tretter, Jones, Andre, et al., 2006), Jones, Taylor, and Broadwell (2009) evaluated the efficacy of teaching about size and scale using body measurements to represent common metric units (e.g., meter, cm, mm). Jones et al. (2009) found teaching using body measures as an effective tool, as students' (N=19) metric length estimations improved significantly following instruction.

While some research has been used to develop pedagogical practices that can more broadly address performance expectations outlined in the NGSS (e.g., Krajcik, Codere, Dahsah, Bayer, & Mun, 2014), two main studies evaluating size and scale instruction have led to the development of learning trajectories or learning progressions for understanding size and scale (e.g., Delgado, 2009; Jones & Taylor, 2009). Jones and Taylor's learning trajectory incorporated both retrospective accounts from experts as well as real-time observations of individuals learning and reasoning about scale, while Delgado's (2009) learning progression incorporated only the latter, because, as he argued, retrospective accounts from experts may or may not accurately recall exactly *how* they acquired a sense of scale or the ways in which in they may have struggled to learn about scale. Both learning

progressions highlighted the fundamental relationship between the acquisition of concepts of absolute size and relative scale (Delgado, 2009; Jones & Taylor, 2009).

Jones and Taylor (2009) used interviews of professionals that use scale in their work as well as prior research with both youth and adults' concepts of size and scale to develop a learning progression categorized by expertise. According Jones and Taylor's progression, novices are developing measurement estimation skills, conceptualizing relative sizes, learning to use measurement tools, and developing a sense of number. Developing learners of scale are able to convert measurements, conceptualize surface area to volume relationships, demonstrate awareness of changing scales, use body measures for length estimations, visualize and understand different types of scales, and develop proportional reasoning and visual spatial skills. Experienced learners of scale demonstrate automaticity and accuracy, can create reliable scales, can relate one scale to another, and can make use of conceptual anchors for estimation tasks (Jones & Taylor, 2009).

Delgado (2009) developed his learning progression based data collected in two studies. The first study involved interviews of students (N=101) in grades 6 through undergraduate, and the second study interviewed a group of the students that participated in the first study and also attended a week-long summer camp that included instruction in size and scale. The students in the second study (N=24) participated in task-based interviews before and after taking part in a nano-themed camp that included instruction on size and scale and these data were primarily used for the development of the learning progression. Evaluation of the instruction indicated that students improved pre to post in the estimation of both absolute scale and relative scale, but the improvement was significant only for absolute

scale ($p < 0.005$, $d = 0.95$) (Delgado, 2009). From a series of data analyses on the patterns in students' reasoning about the size and scale of objects, Delgado (2009) developed a six tiered learning progression which focused on the interconnectedness of four constructs of scale: grouping, ordering, relative, and absolute. In his learning progression, Delgado posited that as individuals gain expertise, they form increased connections between the four constructs. As part of the learning progression, Delgado included characteristics of each level of the progression such as specific landmarks that could be identified, the smallest object or unit that could be identified, logical skills that might be leveraged, strategies that might be employed, and instructional content that would be appropriate (2009). For example, a "level 0" or novice would be unable to make connections related to size and scale between grouping, ordering, relative, or absolute. A level 0 could identify size landmarks only at the human scale and the smallest unit he or she could recall would be a centimeter or an inch. The strategies that a novice could use would include tasks like comparing objects part-to-whole, and suggestions for instruction included introducing the submacroscopic world (e.g., cells, bacteria), transitive properties, and relationships between units such as inches, centimeters and millimeters (Delgado, 2009). In contrast, a "level 5" or expert could make connections between each of the four constructs (grouping, ordering, relative, and absolute). A level 5 could identify size landmarks along a linear number line ranging from microscopic to the Earth, and could identify units as small as nanometers. The strategies that an expert could use include tasks such as proportional reasoning and unit conversions, and suggestions for instruction included a comparison of log and linear scales (Delgado, 2009).

Common limitations across instructional intervention studies included a somewhat short (less than one week) instructional session, self-selection of participants, and lack of a control group by which to compare participants' learning gains (e.g., Delgado, 2009; Jones et al., 2009; Jones et al., 2011). Due to these limitations and mixed results in the efficacy of instruction, there is a need to examine the degree to which exposure to instruction can predict students' concepts of size and scale. This current study differs from prior research involving instructional interventions because the interventions occurred over a relative long (3-year) period, participants included all members of a sixth grade class at a public middle school which limited self-selection, and included a control and experiment group from a population of students at the same school in the same grade level.

Not all factors that may lead to the development of a sense of scale are acquired through learning. This study also examined factors that cannot be controlled by teacher or learner but have been documented as factors that play a role in STEM education such as innate sense of number, gender identity, and racial/ethnic identity.

Innate Systems of Number. Research has shown that individuals with some amount of education represent and process numbers using two systems. The first is the approximate number system (ANS), which processes approximate or imprecise estimations of number (i.e., without counting) (Dehaene, Dehaene-Lambertz, & Cohen, 1998), and the second is the exact number system, which processes exact or precise numbers by way of representative symbols (e.g., number words or quantitative numerals) (Dehaene, 1992). The ANS is distinct from exact number representations because the ANS is present in pre-verbal populations (e.g., human infants (Izard, Sann, Spelke, & Streri, 2009)) and non-verbal populations (e.g.,

rats (Meck, Church, & Gibbon, 1985), monkeys (Piazza & Izard, 2009), and fish (Dadda, Piffer, Agrillo, & Bisazza, 2009), indicating that the ANS is innate and does not require symbolic knowledge or language. The ANS has been described as evolutionarily advantageous because the ability to discriminate quantities of things such as food or predators at a glance could potentially increase the likelihood of survival (Feigenson, Libertus, & Halberda, 2013). Not present at birth, the exact number system is acquired during early childhood alongside the acquisition of language as children map symbolic numbers to representative quantities (Carey, 2009; Le Corre & Carey, 2007). Together, these two systems of number support the complex, culturally based acquisition of mathematics.

A series of studies have examined the degree to which innate intuitions about large numbers and quantities, which are supported solely by the ANS, contribute to concepts of formal mathematics (for a review see Feigenson et al., 2013 and Schneider et al., 2016). Though there is some debate among researchers as to the explanations of why this relationship exists, researchers largely agree that students with greater ANS acuity, as measured through non-symbolic discrimination tasks (e.g., determining which set of different colored dots is of a greater quantity), have higher achievement mathematics (Chen & Li, 2014; Halberda, Mazocco, & Feigenson, 2008; Schneider et al., 2016). Science education researchers, particularly those interested in students' concepts of size and scale, have looked to neuropsychology research on the ANS to help explain students' understanding of size and scale because of the seemingly close relationship between measurement estimation and quantity discrimination (e.g., Delgado et al., 2017; Tretter, Jones, Andre, et al., 2006). More recent research has examined the relationships between students' ANS acuity and students'

accuracy of measurement estimation as well as the relationship between students' ANS acuity and students' science and mathematics achievement (Delgado et al., 2017). Research by Delgado et al. (2017) found a significant positive relationship between students' (N=30) accuracy estimating a meter and students' ANS acuity but found that students' mathematics achievement, science achievement, and accuracy estimating other sizes (e.g., millimeter, centimeter, inch, foot) were not significantly related to students' ANS acuity. Due to the relatively small sample size and mixed results in the relationship between estimating sizes and ANS acuity in the study by Delgado et al. (2017), there is a need for further investigation of this relationship as well as a need to examine ANS acuity as one of many factors that may explain students' concepts of size and scale.

Gender and Racial Identities in Science. Gender and racial discrepancies in science have been well documented in science education research literature and include topics ranging from interest, motivation, career trajectories, achievement, tool use, teacher-student interactions, and student-student interactions, but less is known about these discrepancies related to concepts of size and scale. Research in mathematics education has documented gender and racial differences in students' measurement abilities (e.g., Lubienski, 2003; Lubienski & Shelley, 2003; Thompson & Preston, 2004), but few studies examined these differences in gender and racial groups in the context of science. Mullis et al. (2003) reported that the Trends in Mathematics and Science Study (TIMSS) data showed that students around the world have poor measurement skills, but that U.S. trails the international average in measurement by a greater amount than other mathematics content strands such as algebra or geometry. Lubienski and Shelley (2003) found that on the National Assessment of

Educational Progress (NAEP), students in grade eight demonstrated improvement from 1990 to 2003 in measurement, but this improvement was not among all groups of students. For example, the achievement gap between African American and Caucasian students in measurement on the NAEP increased more than any other mathematical strand from 1990 to 2003. Lubienski and Shelley (2003) also reported that this difference between racial groups transcends socioeconomic levels, as minority students at all levels of socioeconomic status performed lower on measurement items than their non-minority peers. Thompson and Preston (2004) argued that students who lack measurement skills may be less inclined to pursue “measurement-intensive careers” such as those in science, technology, engineering, or mathematics, and may help to explain underrepresentation of minorities in these careers. Thompson and Preston suggested that integrating measurement into science might help to improve measurement education (2004). McGraw, Lubienski, and Strutchens (2006) examined gender differences on the NAEP in mathematics from 1990 to 2003 and found that males slightly outperformed females in each year, particularly in measurement.

Jones and colleagues conducted one of the few studies that included racial differences as well as size and scale (Jones, Tretter, et al., 2007). Jones et al. (2007) investigated differences in African-American and European-American students’ attitudes towards and engagement with nanoscale science instruction. Although the intent of this research was not to examine individual differences in concepts of scale, results demonstrated notable differences African-American and European-American attitudes and perceptions of nanoscale science that may inform future investigations. For example, Jones and colleagues (Jones, Tretter et al., 2007) found that African-American students demonstrated fewer

changes in attitude pre to post instruction than their European-American peers and were significantly more likely to report they felt science involved rote memorization for the purpose of finding a *correct* answer than their European-American peers. Additionally, when writing about their experiences with nanoscale science, African-American students often used formal writing in the third person while European-American students often used the first person (Jones, Tretter, et al., 2007). The authors posited that these differences across groups indicate that African-American students may have different experiences than their European-American peers in science lessons, particular those involving size and scale (Jones, Tretter, et al., 2007). The following chapter outlines the research design and methods employed during this research project.

Methodology

Participants. The population sample consisted of a group of eighth grade students (N=232) who attended a public magnet middle school in the southeastern United States. The school was chosen through convenience sampling due to proximity to the university and existing partnerships with the university research team. The school was considered low-performing and had been designated by the North Carolina report card as a “Grade D” school due to low performance on state standardized tests, called End-of-Grade tests (EOGs) (NC Report Card, 2015). According to the NC Report Card (2015), state and district level proficiency on all EOGs in grades 6 through 8 was reported as 66.1% and 56.3%, respectively, while proficiency at the participating school was reported as 42.5%. Table 1 outlines demographic information for study participants. For the purposes of data analysis,

Non-Caucasian students included students who identified as African-American, Hispanic, Asian, and Other.

Table 1

Participant frequencies by gender and race/ethnic group

Racial/Ethnic Identity	Gender Identity		
	<u>Male</u>	<u>Female</u>	<u>Total</u>
White	72	52	124
Non-white	66	42	108
Total	138	94	232

Note. Non-white participants included those who identified as African American, Hispanic, Asian, and Other.

Study Design. The study utilized a quasi-experimental posttest only control group design (Gall, Gall, & Borg, 2007). The quasi-experimental design was selected due to the following limitations inherent with working in public schools. While students were not selected at random, entire classes were randomly designated as control or experiment. Table 2 shows basic demographic information for the control and experiment groups.

Table 2

Demographic information for control and experiment groups

Gender	Control			Experiment			Total
	<u>White</u>	<u>Non-White</u>	<u>Total</u>	<u>White</u>	<u>Non-White</u>	<u>Total</u>	
Male	33	41	74	39	25	64	138
Female	37	24	61	15	18	33	94
Total	70	65	135	54	43	97	232

Note. Non-white participants included those who identified as African American, Hispanic, Asian, and Other.

Experimental designs are considered the most robust way to establish a cause and effect relationship while controlling for external variables (Gall et al., 2007). The experiment group participated in size and scale instruction of at least 250 minutes or more in grades 6, 7, and 8. Students who were not in the experimental group worked on alternative assignments not related to size and scale but relevant to the grade 8 science curriculum. In grade 8 the control group (n= 97) and the experiment group (n=135) were given post assessments only to measure concepts of scale (SOQ), ANS acuity (Panamath), and science scale capital (ESSS). Basic demographic information was also collected at this time.

Intervention. The students in the experiment group participated in a three-year intervention during which time science educators conducted research-based activities with the students during either science or elective class periods. In each year, students participated

in a series of lessons designed to improve understanding of size and scale, based on prior research in scale as well as proposed learning progressions (Delgado, 2009; Jones & Taylor, 2009). Activities included measurement estimation using body measures, showing the film *Powers of Ten*, metric conversion activities, and interactions with microscopic and nano-sized objects. Table 3 summarizes the activities completed by the experiment group during the study.

Table 3

Descriptions of activities done with the experiment group across all years of the study

Activity	Description
Size Ranges	Students are introduced to the wide range of sizes from the nanoscale to the galactic scale
Metric Body Measures	Students learn how to use their body to estimate the size of objects
Size Scavenger Hunt	Students use their metric body measures to find objects that are 1 cm, 10cm, 1 meter, and 10 meters in length
Powers of 10 (Smaller Units)	Students view the Powers of Ten film and discuss metric units, exponents, and decimal places, specifically for small units (cm, mm, nm)
Zooming with microscopes	Students observe images taken with microscopes to examine the unseen world and discuss how zooming allows us to see things at different scales
Estimating length	Without measuring, students estimate the length of various objects (desk, book, door) using metric units
Powers of 10 (Larger Units)	Students view the Powers of Ten film and discuss metric units, exponents, and decimal places, specifically for large units (m, km)
Your Legs as a Ruler	Students practiced using body measures to pace out distances and use their legs as rulers
Measurement Olympics	Students competed to have the most accurate measurement estimations using their body measures
Metric Buddies	Students explored the metric units and their corresponding exponents
Metric Body Measures	Students reviewed how to use their body to estimate the size of objects and estimated the size of objects using their body rulers

Table 3 Continued

Activity	Description
How Big? How Small?	Students estimated the size of objects in small groups and discussed the correct/incorrect answers
Measuring using Body Measures and Rulers	Students measured objects using both their body measures and metric rulers
Percent Error	Students compared measurements from body rulers to results from metric rulers and calculated percent error for each
Accuracy & Precision	Using measurements from percent error activity, students explored the differences in accuracy and precision
Sketching Models	Students explored scaling strategies and building models to scale
Building Models	Students constructed scaled models first by drawing and then by building them using play-dough
One in a Billion	Students complete a dilution activity to investigate the differences between one in a million and one in a billion. Students examine the relationship between million and billion
How Big is an Atom?	Students explore the atom and how the size of the atom compares to other objects (compound, virus, cell)
How Big is a virus?	Students explore a virus and how the size of a virus compares to other objects (atoms, compounds, cells)
Virus Models	Students build a model virus to scale and discuss the scaling strategies used to generate the model
Geologic Time Scale	Students create geologic timelines and discuss the age of the earth

Instruments.

Size and Scale Instruments. To evaluate students' concepts of scale, students completed the Scale of Objects Questionnaire (SOQ). The SOQ was developed by Tretter, Jones, Andre, et al. (2006) who based this development on earlier work by Trend (2001). The 26-item SOQ indexes students' understanding of absolute scale by asking students to determine if an item falls into one of twelve size categories ranging from "less than one billionth of a meter" to "1 million to 1 billion meters." For example, students were asked to determine the size category for items such as the width of a human hair, to which the correct answer would be "one millionth to one thousandth of a meter." The SOQ was reliable with a Cronbach's Alpha of 0.763 for the sample population.

Students' science scale capital was evaluated using a survey developed for this study called the Experiences with Scale Survey (ESSS) (Appendix A). This 60-item survey asked students to report whether or not they have participated in activities such as estimating how long a trip might take or built a model to scale as well as whether or not they have access to scientific measurement tools at home such as a thermometer or a meter stick. The ESSS was developed based on the ASPIRES Survey (Archer et al., 2012) but with an intentional focus on access to tools. The ESSS was developed by an expert panel of three science education researchers and two science teachers. The ESSS was piloted, revised, piloted a second time, and revised a final time to insure content validity. The ESSS was deemed reliable with a Cronbach's Alpha of 0.931 for the sample population.

Panamath. The Psychophysical Assessment of Numerical Approximation (*Panamath*) instrument was used to index students' ANS acuity (Halberda, Ly, Wilmer,

Naiman, & Germine, 2012). The Panamath test flashes dots of two different colors on a screen and requires the participant determine, without counting, which color dots are more numerous (Halberda et al., 2012). Figure 2 shows an example of dot arrays from the *Panamath* instrument. Developed by Halberda and colleagues, this instrument is widely used by neuropsychologists to index students' ANS acuity for non-symbolic numerical comparison tasks (e.g., Halberda et al., 2008; Libertus, Odic, Feigenson, & Halberda, 2016; Price, Palmer, Battista, & Ansari, 2012) and has demonstrated inter-test reliability across other approximate number system assessments (DeWind & Brannon, 2016). The Panamath was reliable with a Cronbach's Alpha of 0.97 for the sample population.

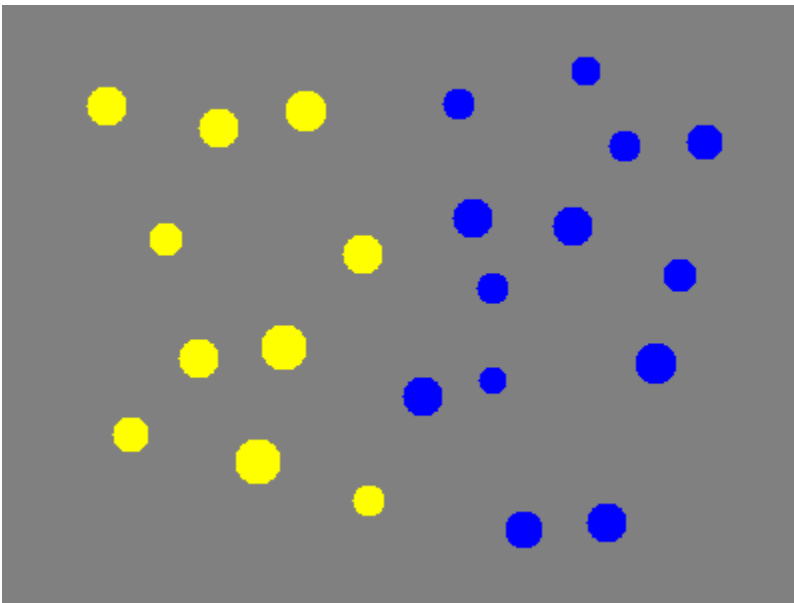


Figure 2. An example from the *Panamath* instrument. The image above would appear approximately 2 microseconds, during which time participants must determine which set of dots (blue or yellow) is of a greater quantity.

Analyses.

Scoring assessments. The following sections describe the processes used to assign student scores for each of the assessments: SOQ, Panamath, and ESSS.

Scoring the SOQ. The SOQ was scored for accuracy, but allowed for partially correct responses. Students earned one point for each test item for selecting the correct size category and half of a point for selecting an adjacent size category (either one power of ten smaller or larger). For example, given the object “length of a football field from goal to goal” a student would earn one full point for selecting the size category “10 meters to 100 meters,” and would earn half of a point for selecting a size category within one power of ten, either smaller (1 meter to 10 meters) or larger (100 meters to 1,000 meters). In total, students could earn a maximum score of 26 points on the SOQ.

Scoring the Panamath. The Panamath, designed to index students’ ANS acuity, was scored for accuracy. While there has been considerable debate about the scoring mechanisms employed for indexing ANS acuity (Price et al., 2012; Libertus et al., 2016), Inglis and Gilmore (2014) found that accuracy (the number of correct responses) was an adequate means of indexing the measure. Students earned one point for selecting the color of dots (blue or yellow) that were more numerous and earned zero points for selecting the incorrect color. Students could earn a maximum score of 60 points on the Panamath.

Scoring the ESSS. Scores on the ESSS were tallied based on number of resources or experiences to which students reported having access. For example, if a student reported having access to a meter stick at home, the student earned one point. In contrast, if a student reported having not ever estimated how long a trip would take, the student earned zero points. In total, a student could earn a maximum score of 60 points on the ESSS.

Missing Data. In total, 312 students consented to participate in the research project in grade 6. Students were removed for one of four reasons: dropping out of the study, failing to complete at least one of the administered assessments, failing to provide necessary demographic information (e.g., gender identity, racial/ethnic identity), or having scores on any of the assessments that were considered substantial outliers (i.e., ± 3 standard deviations). Across the three years of the study, 79 students were removed, leaving 232 total students in the analyses.

Multiple Imputation. Despite the removal of 79 students, some of the remaining students had missing information because they missed one or more assessments during the study. To address missing data, data were imputed through a multiple imputation using SPSS software. The idea of expanding data analyses beyond “complete-case” datasets was first introduced by Rubin (1976) who argued that complete-case analysis (i.e., including only cases for which the researcher has a complete dataset) limits the scope of research. One method of including cases with missing data into the analysis is to complete a multiple imputation of the data. Royston (2004) described data analysis with multiple imputations as creating a relatively small number of copies of the data (in this current study 5 copies were generated), and each copy “has the missing values suitably imputed,” (p. 228) and the newly completed datasets are analyzed independently. Royston explained further “estimates of parameters of interest are averaged across the m copies to give a single estimate” (p. 228).

Prior to conducting the multiple imputations, data were analyzed for patterns of missing data and degree of missing data. Figure 3 summarizes missing values (i.e., individual scores on each assessment) and cases (i.e., individual student participants in the study) in the

original data. As shown in Figure 3, 2.73 % (N=232) of the values in the dataset were missing and required imputation and 15.95 % of the cases (i.e., individuals) required imputation of one or more values. Patterns of missing data were generated in SPSS and are shown in Figure 4, demonstrating no major patterns in missing data. Following the analysis of patterns of missing data, a multiple imputation was performed in SPSS, which generated 5 distinct datasets. The new datasets included 232 total cases, all of which had no missing information.

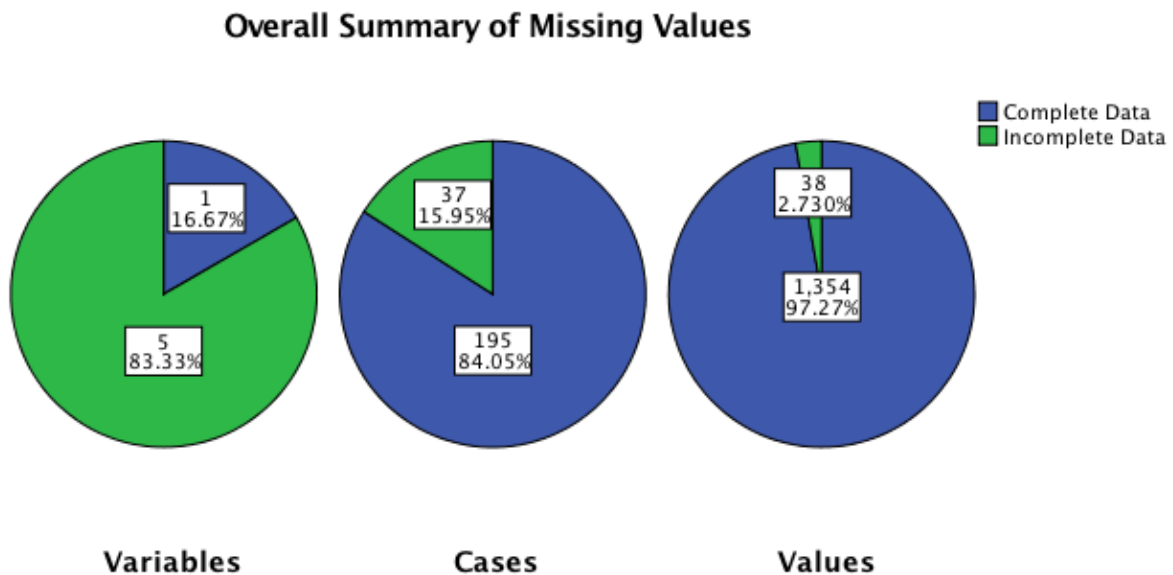


Figure 3. *Summary of missing values prior to imputation. Variables include the factors examined in the study (gender, race/ethnicity, instruction, ANS acuity, SOQ scores, and ESSS scores) (n=6). Cases refer to individual students in the study (n=232). Values refer to the scores for each student on each assessment (n=1392)*

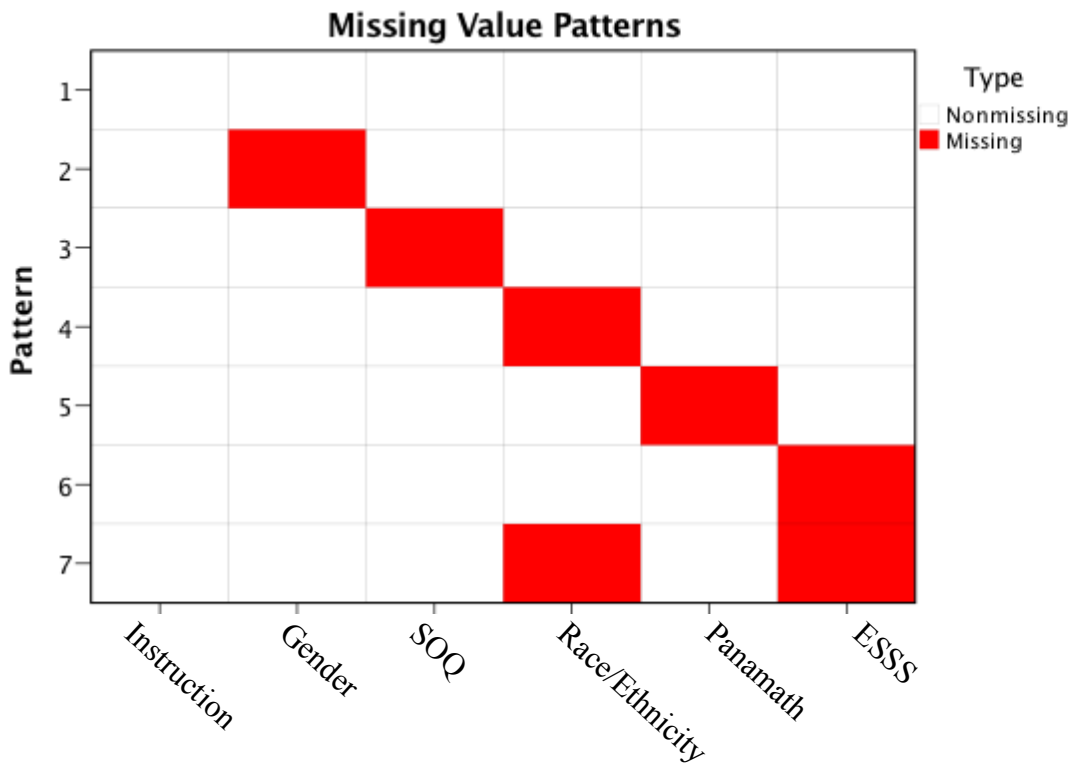


Figure 4. *Patterns of missing values produced in SPSS. A visual inspection of the chart above indicates there are no clear patterns of missing values across variables.*

Multiple Regression. Once missing values were imputed, a linear multiple regression was conducted to predict students’ concepts of scale, as measured by performance on the SOQ, from a series of factors such as ANS acuity, science scale capital, gender identity, and racial/ethnic identity. Results from the multiple regression demonstrate the proportion of the variation in students’ scores on the SOQ that can be explained by the independent variables (i.e., ANS acuity, science scale capital, gender, or race/ethnicity as well as determine how much students’ scores on the SOQ might change as a result of one unit change in the independent variable.

Prior to interpretation of the results, the output of the multiple regression was analyzed to ensure that the data met the assumptions necessary for reliable interpretation. The assumption of linearity was met as assessed by partial regression plots and a plot of studentized residuals against the predicted values. The assumption of independence of residuals was met as assessed by a Durbin-Watson statistic of 1.820 for the original data. The assumption of homoscedasticity was met, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Another assumption is that there should be no evidence of multicollinearity, which occurs when a strong correlation exists between two or more independent variables. None of the independent variables had correlations greater than 0.7, which served as one indication of the lack of multicollinearity. Other indications of no multicollinearity included a lack of tolerance values less than 0.1, no leverage values greater than 0.2, and no values for Cook's distance above 1. The assumption of normality was met, as assessed by P-P Plots of both the original and imputed data. Once each of the assumptions was deemed as met, the analyses of the multiple regression were conducted.

The null hypothesis of the multiple regression was that the multiple correlation coefficient (R) is equal to zero. The alternative hypothesis is that R is not equal to zero. The following section reports on the results of the study, and the regression equation can be expressed as the following:

$$SOQ = b_0 + (b_1 * ANS) + (b_2 * instruction) + (b_3 * science\ scale\ capital) + (b_4 * gender) + (b_5 * race)$$

Results

The results of the multiple regression model for the imputed data are reported in the sections that follow. Pooled values, rather than those generated from the original data or individual imputations, were used because these values took into account the models built from each of the five data imputations. When pooled data were not available due to limitations in the SPSS software (Van Ginkel & Kroonenberg, 2014), either averages or ranges were reported for the models built from each of the five imputations. The multiple regression model was used to determine the proportion of the variation in students' scores on the SOQ that can be explained by the independent variables (i.e., ANS acuity, science scale capital, instruction, gender, or race/ethnicity as well as determine how much students' scores on the SOQ might change as a result of one unit change in the independent variable, which is of particular importance for the nominal independent variables (i.e., instruction, gender, or race/ethnicity). The multiple correlation coefficient (R) is the Pearson correlation coefficient demonstrating the relationship between the scores predicted by the regression model and the actual values of the students' SOQ scores. The strength of this relationship is somewhat strong for the imputed data ($0.607 < R < 0.629$) indicating a moderate linear association. The coefficient of determination (R^2) is a measure of the proportion of variance in the dependent variable (SOQ performance) that is explained by the independent variables (science scale capital, ANS acuity, instruction, gender, or race/ethnicity). R^2 for the overall model ranged from 36.8% to 39.5% with an adjusted R^2 range of 35.4% to 38.2%, a somewhat small but acceptable effect size according to Cohen (1988). The multiple regression model statistically and significantly predicted students' performance on the SOQ, $F(5, 226) = 27.63, p < 0.0005$,

adj. $R^2 = 0.3674$ (values for F and adjusted R^2 are reported as averages from all five imputations). Only the variables of science scale capital, instruction, and race/ethnicity added statistically and significantly to the prediction, $p < 0.05$. Regression coefficients and standard errors for the pooled data and imputed data can be found in Tables 4 and 5, respectively.

Table 4.

Regression coefficients and standard errors for variables in the pooled data. Standardized coefficients (β) were not available using SPSS software.

Variables	B	SE_B
Intercept	6.27	1.95
Instruction	1.65*	0.62
Gender	0.41	0.64
Race	5.45*	0.67
Science Scale Capital	0.09*	0.02
ANS Acuity	0.04	0.04

Note. * $p < 0.001$. B = unstandardized regression coefficient; SE_B = standard error of the coefficient

Table 5.

Regression coefficients and standard errors for variables in each of the five imputations

<u>Variables</u>	Imputation														
	1			2			3			4			5		
	<u>B</u>	<u>SE_B</u>	<u>β</u>	<u>B</u>	<u>SE_B</u>	<u>β</u>	<u>B</u>	<u>SE_B</u>	<u>β</u>	<u>B</u>	<u>SE_B</u>	<u>β</u>	<u>B</u>	<u>SE_B</u>	<u>β</u>
Intercept	5.94	1.89		6.59	1.92		6.81	1.92		6.04	1.87		6.00	1.90	
Instruction	1.69	0.61	0.15*	1.72	0.62	0.15*	1.72	0.62	0.15*	1.68	0.60	0.15*	1.47	0.61	0.13*
Gender	0.28	0.62	0.02	0.23	0.63	0.02	0.45	0.63	0.04	0.48	0.61	0.04	0.62	0.62	0.05
Race	5.68	0.64	0.49*	5.39	0.65	0.47*	5.56	0.64	0.49*	5.30	0.64	0.47*	5.35	0.64	0.47*
Science Scale															
Capital	0.09	0.02	0.22*	0.09	0.02	0.23*	0.08	0.02	0.21*	0.10	0.02	0.26*	0.09	0.02	0.24*
ANS Acuity	0.05	0.04	0.06	0.04	0.04	0.05	0.04	0.04	0.05	0.04	0.04	0.05	0.05	0.04	0.06

Note. * $p < 0.001$. B = unstandardized regression coefficient; SE_B = standard error of the coefficient; β = standardized coefficient

The values of the slope coefficients indicate the degree to which changes in factors such as instruction (from control to experiment), gender (from female to male), and race (from non-Caucasian to Caucasian) can predict student performance on the SOQ. Students in the experiment group were predicted in the model to perform 1.65 points higher on the SOQ and, though not significant, males were predicted in the model to perform 0.41 points higher on the SOQ. There was also a significant difference between white and non-white students predicted performance on the SOQ in the model, with Caucasian students predicted to outperform their non-Caucasian peers on the SOQ by 5.45 points. For continuous variables, the model also predicted a positive slope, meaning that higher scale capital and more acute ANS predicted slightly better performance on the SOQ, but only scale capital was statistically significant.

A Kendall's tau-b correlation was run to determine the relationship between students' ANS acuity and students' performance on the SOQ for participants ($n = 215$). There was a weak, positive association between students' ANS acuity and students' performance on the SOQ, which was statistically significant, $\tau_b = .103, p = .030$.

Discussion

Past studies on students' concepts of scale have revealed that factors such as proportional reasoning (Taylor & Jones, 2007; Taylor, 2009), logical thinking (Taylor, 2009), age (Tretter, Jones, Andre, et al., 2006), and expertise (Jones & Taylor, 2009; Tretter, Jones, Andre, et al., 2006) are related to students' understanding of size and scale. The present study showed that factors such as science scale capital and students' racial/ethnic identities can

play a statistically significant role in predicting students' concepts of scale. Not statistically significant from the model in this study were factors of gender and ANS acuity. First, a brief discussion of non-significant factors in the model is included followed by a discussion of scale capital, exposure to instruction, and racial/ethnic identities, all of which were significant in the regression model. Lastly, implications for educational practice as well as future areas of research are addressed.

Non-significant Factors.

Gender. The variables of gender and ANS acuity had no statistical significance in the regression model. Very little is known about the degree to which these factors may influence students' concepts of size and scale, particularly within the context of science. Recent research has examined differences between gender groups as it relates to scientific understanding of spatial relationships (e.g., lunar models (Wilhelm, 2009). While understanding of spatial relationships is not synonymous with concepts of size and scale, Jones and colleagues (2011) found a statistically significant relationship between students' understanding of scale related to zooming and students' spatial reasoning. Research in understanding of spatial concepts has demonstrated differences between genders, with males historically outperforming their female peers (Kerns & Berenbaum, 1991; Silverman, Choi, & Peters, 2007). However, Wilhelm and colleagues (Wilhelm, 2009; Wilhelm, Jackson, Sullivan, & Wilhelm, 2013) found that despite lower initial scores in spatial reasoning, females demonstrated greater gains in experimental studies in which females and males both received intervention related to spatial reasoning. Though there was no significant influence of gender in predicting students' performance on the SOQ in grade 8, further research is

needed to determine if gains from grade 6 to grade 8 differed across gender groups or if there were differences in female performance on scale assessments across experiment and control groups.

ANS acuity. The innate ability to quickly and approximately process quantity, as supported by an acute ANS, has demonstrated a weak, positive but significant relationship with mathematical achievement (Halberda et al., 2008; Libertus et al., 2013). While some science education researchers interested in concepts of size and scale have posited that the ability to estimate size and understand scale may leverage the ANS (e.g., Delgado et al., 2017; Tretter, Jones, Andre et al., 2007), no research to date has been able to successfully prove this relationship exists. Although students' ANS acuity did not serve as a statistically significant factor in predicting students' performance on the SOQ in the regression model, which took into account all other factors, there was a statistically significant relationship, when considering only students' ANS acuity and students' concepts of absolute scale. These findings align with some research on the relationship between students' ANS acuity and academic achievement that considered additional factors such as inhibitory control or socioeconomic status. For example, Fuhs and McNeil (2013) reported that when taking into account students' inhibitory control (the ability to ignore unneeded details), ANS acuity failed to demonstrate a significant relationship with mathematical achievement. One explanation may be that the weak but significant positive correlation between ANS acuity and performance on the SOQ is that ANS acuity supports mathematical ability which is the true underlying factor of influence in students' concepts of size and scale as measured by the SOQ. There is a need to investigate the degree to which ANS acuity or concepts of size and

scale can predict students' performance on standardized achievement tests in mathematics and science.

Significant Factors. Students' reported science scale capital as measured by the ESSS, students' exposure to size and scale instruction, and students' reported racial/ethnic identities all contributed significantly to the regression model for this study. Both science scale capital and racial/ethnic identities have underlying socioeconomic ties, which call for an examination of students' access to resources through a social justice framework. Adams and Bell (2016) refer to social justice as more than a goal; the authors explained that social justice also encompasses the sequence of actions that are needed to attain the goal. The objective of social justice is to create an environment that is both equitable and sustainable (Adams & Bell, 2016).

Science scale capital. Students' reported science scale capital, as measured by the ESSS, was statistically significant as a predictor of students' performance on the SOQ. The importance of out-of-school experiences and resources are supported by prior research which found that adults reported out-of-school experiences as some of the most influential in shaping their understanding of size and scale (Jones & Taylor, 2009). These results also confirm the findings of Archer et al. (2012) that there are individual differences in access to science experiences and science support networks, and while Archer and colleagues focused on students' career aspirations, this study demonstrates that access to science scale capital can also have a significant and positive effect on students' concepts of size and scale. This is not surprising, as previous studies have found similar results through adults' retrospective accounts of what led to their understanding of size and scale (e.g., Tretter, Jones, Andre, et

al., 2006; Jones & Taylor, 2009). This study was different in that it examined the degree to which these reported science scale capital experiences and resources could predict students' performance on size and scale assessments in real time. Appendix B reports percentages of students who reported having access outside of school to specific types of tools and experiences. The degree to which specific experiences and tools may be related to students' concepts of size and scale warrants further investigation.

Given the emphasis being placed on crosscutting concepts such as size and scale in science, the results of this study raise questions about access to science learning tools and science learning experiences for students from disadvantaged homes. If specific resources or experiences demonstrate relationships with students' concepts of size and scale, it could have beneficial implications for the experiences and resources teachers and educational stakeholders could provide to students. There is need for further investigation of the differences between students' access to science scale capital across gender and race/ethnic groups, specifically if there are differences in the specific tools and experiences to which each gender and/or racial/ethnic group reports having access. This is important in light of Archer and colleagues' research that showed that unequal distributions of both science capital and family science habitus demonstrated associations with uneven patterns in students' science career aspirations which existed across class and racial groups, a discrepancy that may help to explain disproportionate representation of gender and racial/ethnic minorities in STEM careers.

Size and Scale Instruction. Students' participation in the intervention (i.e., experimental group) played a significant role in predicting students' concepts of scale as

measured by performance on the SOQ. This is supported by other research, which found significant improvement in students' concepts of scale following instruction (e.g., Jones, Taylor, & Broadwell, 2009; Delgado, 2009). This is promising as it indicates that instruction can support students' concepts of size and scale and that teaching about size and scale may contribute to overall student learning in science and mathematics. Further studies can document the specific size ranges (e.g., nanoscale or macroscale) in which students made the most robust gains in understanding and the specific activities (see Tables 4 and 5) that were the most effective in teaching about concepts of size and scale for specific groups of students (e.g., gender or racial/ethnic). Given the influence of both science scale capital and racial/ethnic identities on predicting students' concepts, factors that are largely out of teachers' control, it is promising that instruction, a factor over which teachers do have control, can affect change in students' concepts of size and scale.

Racial/Ethnic Groups. The results of this study indicated that Caucasian students significantly outperformed their non-Caucasian peers on the SOQ, when taking into account other factors such as gender, ANS acuity, exposure to instruction on size and scale. This observed difference across racial/ethnic groups likely has socioeconomic underpinnings and may be closely tied to students' experiences with size and scale outside of school, thus it may be the racially-linked science scale capital, rather than race alone which is the underlying factor. This is supported by Archer et al. (2012) who found differences not only across racial/ethnic groups of more than 9,000 participants but also in "class" of socioeconomic status. Another explanation may be that students from non-Caucasian backgrounds have different experiences in science class, particularly those related to size and scale instruction,

than their Caucasian peers. This was supported by Jones, Tretter et al. (2007) who found that African-American students demonstrated fewer changes in attitude pre to post instruction on science related to the nanoscale than their European-American peers. Though not directly related to size and scale instruction, learning about the unseen world (i.e., objects at the microscale or nanoscale) is one component of size and scale learning progressions (Delgado, 2009; Jones & Taylor, 2009). Furthermore, African-American students in Jones' and colleagues study were less likely to write about their experiences during the nano science activities in the first person than their European-American peers, indicating that students who identify as African-American may have altogether different experiences with science both in and out of school as it relates to exploring science at the nanoscale. These differences in experiences may lead to discrepancies in concepts of size and scale. There is need to examine the context and the degree to which students from different racial/ethnic backgrounds experience science both in and out of school.

Future Research and Implications for Education

This study brings to light several new areas of research that warrant a closer examination. There is a need to examine the degree to which specific science experiences and specific access to measurement tools may play a role in how students reason about scale and whether there is alignment between experiences with specific sizes and scales and accuracy of estimation at those sizes and scales. Furthermore, it is of interest to examine the degree to which specific aspects of science scale capital, or science capital more generally, can predict students' science career aspirations. While science self-efficacy or perceived science ability serves as one factor in determining students' career choice (Lent, Brown, &

Hackett, 1994), it may be beneficial to the science education research community to conduct a controlled experiment in which students with high science interest gain access to science capital to which they had no prior access such as technology (i.e., iPads or tablet computers), museum programs, and adults with STEM careers who would be willing to form mentoring relationships. Such an investigation may shed light on the importance of access to both resources and experiences for students from groups that are traditionally underrepresented in STEM.

Crosscutting concepts such as size and scale have been argued to have associations with deep learning of science content (NGSS Lead States, 2013) as well as science and mathematics achievement (Chesnutt et al., in preparation). As such, there is a growing need to apply the factors that may predict students' concepts of size and scale to everyday pedagogy. For example, teachers should leverage students' prior experiences (Delgado, 2009; Tretter, Jones, Andre, et al., 2006; Jones & Taylor, 2009) as well as aspects of science scale capital when teaching about scale in the context of science. To do this, teachers could inventory students' access to measurement tools and experiences related to measurement, estimation, conversion of units, or informal science learning environments to better understand the resources and experiences teachers could provide to their students. To ensure that all students have equitable access to learning opportunities in science, teachers should also consider that students of varying backgrounds might have different experiences engaging in science content, particularly as it relates to crosscutting concepts such as size and scale (Jones, Tretter, et al., 2007). Furthermore, teachers should be encouraged to adopt

research-based strategies for teaching about scale, and science educators should continue to develop and share lessons and activities that might support concepts of size and scale.

References

- Albarracín, L., & Gorgorió, N. (2014). Devising a plan to solve fermi problems involving large numbers. *Educational Studies in Mathematics*, 86(1), 79-96. doi:10.1007/s10649-013-9528-9
- American Association for the Advancement of Science. (1994). *Benchmarks for science literacy* Oxford University Press.
- Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science aspirations, capital, and family habitus: How families shape children's engagement and identification with science. *American Educational Research Journal*, 49(5), 881. doi:10.3102/0002831211433290
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 42(1), 189-201. doi:10.1037/0012-1649.41.6.189
- Bourdieu, P. (1977). *Outline of a theory of practice* Cambridge University Press.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste* Harvard University Press.
- Bybee, R. (2015). Scientific literacy. In R. Gunstone (Ed.), *Encyclopedia of science education* (pp. 944-947) Springer Netherlands. Retrieved from http://link.springer.com/referenceworkentry/10.1007/978-94-007-2150-0_178

- Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, 133(1), 59-68.
doi:10.1162/001152604772746701
- Carey, S. (2009). *The origin of concepts*. Oxford University Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edition.
Erlbaum, Hillsdale, NJ.
- Cheek, K. A. (2013). How geoscience novices reason about temporal duration: The role of spatial thinking and large numbers. *Journal of Geoscience Education*, 61(3), 334-348.
doi:10.5408/12-365.1
- Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, 148, 163-172.
- Coleman. (1981). *Public and private schools. an analysis of high school and beyond: A national longitudinal study for the 1980's.* Washington, D.C.: National Center for Educational Statistics.
- Creswell, J. W. (2014). *Research design : Qualitative, quantitative, and mixed methods approaches* (4th ed. ed.). Thousand Oaks: SAGE Publications,. Retrieved from <http://catalog.lib.ncsu.edu/record/NCSU2765361>
- Dadda, M., Piffer, L., Agrillo, C., & Bisazza, A. (2009). Spontaneous number representation in mosquitofish. *Cognition*, 112(2), 343-348.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1), 1-42.

- Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in neurosciences*, 21(8), 355-361.
- Deboer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, 37, 582-601. doi:AID-TEA5>3.0.CO;2-L
- Delgado, C. (2009). Development of a Research-Based Learning Progression for Middle School Through Undergraduate Students' Conceptual Understanding of Size and Scale (Doctoral dissertation), The University of Michigan, Ann Arbor, MI.
- Delgado, C. (2013). 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. doi:10.1080/09500693.2013.779761
- Delgado, C. (2014). Collective landmarks for deep time: A new tool for evolution education. *Journal of Biological Education*, 48(3), 133-141. doi:10.1080/00219266.2013.849280
- Delgado, C., & Lucero, M. M. (2015). Scale construction for graphing: An investigation of students' resources. *Journal of Research in Science Teaching*, 52(5), 633-658. doi:10.1002/tea.21205
- Delgado, C., Jones, M. G., You, H. S., Robertson, L., Chesnutt, K., & Halberda, J. (2017). Scale and the Evolutionarily-Based Approximate Number System: An Exploratory Study. *International Journal of Science Education*, 1-17.

- DeWind, N. K., & Brannon, E. M. (2016). Significant Inter-Test Reliability across Approximate Number System Assessments. *Frontiers in Psychology, 7*, 310. <http://doi.org/10.3389/fpsyg.2016.00310>
- Falk, J. H., & Dierking, L. D. (2010). The 95 percent solution. *American Scientist, 98*(6), 486-493. Retrieved from <http://search.proquest.com/docview/847469973/abstract/87DDAE4A07BC4D56PQ/1>
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational Research: An Introduction* (8th ed.). New York, NY: Pearson.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1), 43-74.
- Golledge, R. G., Gale, N., Pellegrino, J. W., & Doherty, S. (1992). Spatial knowledge acquisition by children: Route learning and relational distances. *Annals of the Association of American Geographers, 82*(2), 223-244. doi:10.1111/j.1467-8306.1992.tb01906.x
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences, 109*(28), 11116-11120.
- Halberda, J., Mazocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665-668.
- Hegarty, M. (2014). Spatial thinking in undergraduate science education. *Spatial Cognition & Computation, 14*(2), 142-167. doi:10.1080/13875868.2014.889696

- Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*(2), 151-176.
doi:10.1016/j.intell.2005.09.005
- Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147-155.
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, *106*(25), 10382-10385.
- Jones, M. G., Taylor, A., & Broadwell, B. (2009). Estimating linear size and scale: Body rulers. *International Journal of Science Education*, *31*(11), 1495-1509.
doi:10.1080/09500690802101976
- Jones, M. G., Forrester, J. H., Robertson, L. E., Gardner, G. E., & Taylor, A. R. (2012). Accuracy of estimations of measurements by students with visual impairments. *Journal of Visual Impairment & Blindness*, *106*(6), 351-355.
- Jones, M. G., Gardner, G., Taylor, A. R., Wiebe, E., & Forrester, J. (2011). Conceptualizing magnification and scale: The roles of spatial visualization and logical thinking. *Research in Science Education*, *41*(3), 357-368. doi:10.1007/s11165-010-9169-2
- Jones, M. G., Paechter, M., Gardner, G., Yen, I., Taylor, A., & Tretter, T. (2010, June). Teachers' concepts of spatial scale: an intercultural comparison between Austrian, Taiwanese, and US-American teachers. In *Proceedings of the 9th International*

Conference of the Learning Sciences-Volume 2 (pp. 284-285). International Society of the Learning Sciences.

Jones, M. G., & Rua, M. J. (2008). Conceptual representations of flu and microbial illness held by students, teachers, and medical professionals. *School Science and Mathematics, 108*(6), 263-278. doi:10.1111/j.1949-8594.2008.tb17836.x

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

Jones, M. G., Taylor, A. R., & Broadwell, B. (2009). Concepts of scale held by students with visual impairment. *Journal of Research in Science Teaching, 46*(5), 506-519. doi:10.1002/tea.20277

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. doi:10.1007/s10956-006-9034-2

Jones, M. G., Tretter, T., Paechter, M., Kubasko, D., Bokinsky, A., Andre, T., & Negishi, A. (2007). Differences in African-American and European-American students' engagement with nanotechnology experiences: Perceptual position or assessment artifact? *Journal of Research in Science Teaching, 44*(6), 787-799. doi:10.1002/tea.20168

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. doi:10.1080/09500690701416624

- Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. *Review of Educational Research*, 68(4), 413-449. doi:10.2307/1170734
- Kerns, K. and Berenbaum, S. 1991. Sex differences in spatial ability in children. *Behavior Genetics*, 21: 383–396.
- Lamon, S. (1993). Ratio and proportion: connecting content and children's thinking. *Journal for Research in Mathematics Education*, 24(1), 41-61.
- Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105(2), 395-438.
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of vocational behavior*, 45(1), 79-122.
- Liben, L. S., Kastens, K. A., & Christensen, A. E. (2011). Spatial foundations of science education: The illustrative case of instruction on introductory geological concepts. *Cognition and Instruction*, 29(1), 45-87. doi:10.1080/07370008.2010.533596
- Liben, L. S., Myers, L. J., Christensen, A. E., & Bower, C. A. (2013). Environmental-scale map use in middle childhood: Links to spatial skills, strategies, and gender. *Child Development*, 84(6), 2047-2063. doi:10.1111/cdev.12090

- Libertus, M. E., Odic, D., Feigenson, L., & Halberda, J. (2016). The precision of mapping between number words and the approximate number system predicts children's formal math abilities. *Journal of Experimental Child Psychology*, 150, 207-226.
- Lubienski, Sarah T. "Is Our Teaching Measuring Up? Race-, SES-, and Gender-Related Gaps in Measurement Achievement." In Clements, D. H. & Bright, G. W. (Eds.), *Learning and Teaching Measurement, 2003 Yearbook of the National Council of Teachers of Mathematics*, (282-92). Reston, Va.: NCTM.
- Lubienski, S. T., & Shelley, M. C. (2003). A Closer Look at US Mathematics Instruction and Achievement: Examinations of Race and SES in a Decade of NAEP Data.
- Magana, A. J., Brophy, S. P., & Bryan, L. A. (2012). An integrated knowledge framework to characterize and scaffold size and scale cognition (FS2C). *International Journal of Science Education*, 34(14), 2181-2203. doi:10.1080/09500693.2012.715316
- McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education*, 129-150.
- Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 11(4), 591.
- Mullis, I. V., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, E. J., Chrostowski, S. J., & O'Connor, K. M. (2003). *TIMSS Trends in mathematics and*

science study: Assessment frameworks and specifications 2003. International Association for the Evaluation of Educational Achievement.

National Research Council. (1996). *National science education standards*. [S.l.]: National Academy Press.

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

NC Report Card. (2015). Retrieved from <https://ncreportcards.ondemand.sas.com/landing.html>

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, US: National Academies Press. Retrieved from <http://site.ebrary.com/lib/ncsu/docDetail.action?docID=10863742&ppg=1>

Piaget, J. & Inhelder, B. (2013). *Mental Imagery in the Child: Selected Works* (Vol. 6) New York, NY: Routledge

Piaget, J., Inhelder, B., & Szeminska, A. (Eds.). (1960). *The child's conception of geometry*. New York: Basic Books. Retrieved from http://buprimo.hosted.exlibrisgroup.com/primo_library/libweb/action/dlSearch.do?institution=BOSU&vid=BU&query=any,contains,99195692400001161

Piazza, M., & Izard, V. (2009). How humans count: numerosity and the parietal cortex. *The Neuroscientist*, 15(3), 261-273.

- Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta psychologica*, 140(1), 50-57.
- Resnick, I., Davatzes, A., Newcombe, N. S., & Shipley, T. F. (2016). Using relational reasoning to learn about scientific phenomena at unfamiliar scales. *Educational Psychology Review*, 29(1), 11-25. doi:10.1007/s10648-016-9371-5
- Royston, P. (2004). Multiple imputation of missing values. *Stata journal*, 4(3), 227-41.
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental science*. Advance online publication: doi: 10.1111/desc.12372
- Silverman, I., Choi, J. and Peters, M. 2007. The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Archives of Sexual Behavior*, 36: 261–268.
- Taylor, A. R., & Jones, M. G. (2013). Students' and teachers' application of surface area to volume relationships. *Research in Science Education*, 43(1), 395-411.
- Taylor, A., & Jones, G. (2009). Proportional reasoning ability and concepts of scale: Surface area to volume relationships in science. *International Journal of Science Education*, 31(9), 1231-1247. doi:10.1080/09500690802017545
- Thompson, T. D., & Preston, R. V. (2004). Measurement in the Middle Grades: Insights from NAEP and TIMSS. *Mathematics teaching in the Middle School*, 9(9), 514-519.

- 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(2), 191-221. doi:AID-TEA1003>3.0.CO;2-C
- Tretter, T. R., Jones, M. G., Andre, T., Negishi, A., & Minogue, J. (2006). Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching*, 43(3), 282-319.
doi:10.1002/tea.20123
- Tretter, T. R., Jones, M. G., & Minogue, J. (2006). Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching*, 43(10), 1061-1085. doi:10.1002/tea.20155
- Wilhelm, J. 2009. Gender differences in lunar-related scientific and mathematical understandings. *International Journal of Science Education*, 31: 2105–2122.
- Wilhelm, J., Jackson, C., Sullivan, A., & Wilhelm, R. (2013). Examining differences between preteen groups' spatial-scientific understandings: A quasi-experimental study. *The Journal of Educational Research*, 106(5), 337-351.

Appendices

Appendix A: Experiences with Size and Scale Survey (ESSS)

Part I. Science, Size and Scale Outside of School

1. What do you think you will do for a job or career when you're older?
2. What do you want to do for a job or career when you're older?
3. Do you know anyone that uses size, scale, or measurement in his or her work?
4. Have you ever talked about the sizes of things or distance between places with your friends or family?
5. Do you want to study science in high school or college?
6. Are any of your friends or classmates really into science?

Part II. Tools of Measurement

I. Do you own or have you used any of the following at home? (select one or both)

• map	own	used	never owned or used
• ruler	own	used	never owned or used
• yard/meter stick	own	used	never owned or used
• measuring tape	own	used	never owned or used
• weight scale	own	used	never owned or used
• compass	own	used	never owned or used
• GPS	own	used	never owned or used
• measuring cups	own	used	never owned or used
• weight (bathroom) scale	own	used	never owned or used
• timer	own	used	never owned or used
• thermometer	own	used	never owned or used
• health monitor (blood pressure)	own	used	never owned or used
• chemistry or science kit	own	used	never owned or used
• Lincoln Logs	own	used	never owned or used
• Legos (children)	own	used	never owned or used
• building blocks	own	used	never owned or used

Part III. Experiences with Size and Scale

Have you ever done any of the following? (Indicate those you participated in with a check).

- ✓ Measured or timed swimming/running/obstacle course
- ✓ Measured laps while swimming/running
- ✓ Built a model to scale (E.g. airplane, boat, house, diorama for a class project)
- ✓ Measured or cut wood or fabric
- ✓ Estimated how long a trip would last
- ✓ Estimated how long it would take to walk, run, or ride a certain distance
- ✓ Estimated how much birdseed would fit in a container
- ✓ Estimated how far you rode, swam, ran
- ✓ Estimated the depth of swimming pool
- ✓ Estimated the amount of paint needed to paint something
- ✓ Estimated the amount of materials needed to build something

- ✓ Estimated the amount of food a pet needs
- ✓ Estimated the amount of food you would eat
- ✓ Estimated the amount of clothes or items you could fit in a suitcase
- ✓ Estimated how much data you've used on the computer/phone (megabytes, gigabytes, etc.)
- ✓ Measured temperature using a thermometer
- ✓ Measured how long a trip would last
- ✓ Measured the depth of swimming pool
- ✓ Looked at objects through a microscope/telescope
- ✓ Looked at objects using binoculars
- ✓ Looked at objects through a magnifying glass
- ✓ Gone to a museum exhibit that showed you the size of small objects
- ✓ Gone to a museum exhibit that showed you the size of large objects
- ✓ Seen a show at a planetarium
- ✓ Converted between kilometers and miles
- ✓ Converted between degrees Fahrenheit and degrees Celsius
- ✓ Converted between meters and yards
- ✓ Converted between inches and centimeters
- ✓ Converted between liters and gallons
- ✓ Converted between megabytes, kilobytes, gigabytes, or terabytes
- ✓ Used your body to measure something
- ✓ Changed a recipe to serve more or fewer people
- ✓ Designed or planned a space to put a garden
- ✓ Gardened with your family to grow fruits or vegetables
- ✓ Estimated how long it would take a plant to grow or produce fruit
- ✓ Estimated how much fruit or how many vegetables you would get from a plant
- ✓ Looked up the age of a rock or fossil
- ✓ Looked up the distance between Earth and the Moon
- ✓ Looked up the distance between Earth and the Sun
- ✓ Looked at something using satellite images (such as Google Earth)
- ✓ Created a map to give someone else
- ✓ Watched science movies with your family
- ✓ Gone for a hike outside with your family
- ✓ Gone geocaching with your family or friends
- ✓ Participated in science clubs (afterschool/lunchtime)
- ✓ Participated in activities related to other subjects (such as English or history) outside of school

Appendix B: Tool Use Access by Percentage on the ESSS

Table B1: Percentage of students that reported having access to the tools listed in the survey. Not all students responded to every question, so the sample size varies for each tool.

Tool	Percentage of Students that Report Having Access
Ruler (n=189)	100.00%
Timer (n=186)	97.30%
GPS (n=189)	96.80%
Measuring Tape (n=188)	96.30%
Measuring Cups (n=187)	95.20%
Thermometer (n=185)	94.60%
Legos (n=185)	93.00%
Meter stick (n=189)	92.60%
Map (n=189)	88.90%
Weight Scale (n=188)	86.70%
Compass (n=188)	76.10%
Building Blocks (n=186)	75.80%
Health Monitor (n=186)	55.40%
Chemistry or Science Kit (n=184)	41.90%

Conclusion

As states continue to adopt the Next Generation Science Standards (NGSS), there will be a growing need to teach science not only across science disciplines but also across all curricula. One approach to achieving this goal is to teach science content using broad themes or crosscutting concepts. Such concepts, such as scale, proportion, and quantity allow science to be conveyed to students as a tightly interconnected series of phenomena in which concepts in physics can help to explain chemistry concepts, which help to explain principles in biology. For example, forces and motion from physics help to explain molecular motion in chemistry that can help to explain diffusion in biology. The research studies outlined in this dissertation indicated that concepts of size and scale are closely related to student achievement in both science and mathematics. Furthermore, factors such as racial/ethnic identity, exposure to size and scale instruction, and experiences with resources and tools outside of school support students' concepts of scale. A model for explaining the factors of influence is included in Figure 1. In this model, learning about size and scale is situated within one's culture and innate sense of number. Within the context of culture and number sense are factors of instruction, race/ethnic identity, and science scale capital. All of these factors (instruction, race/ethnicity, and science scale capital) interact. For example, one's race/ethnicity may play a role in the degree to which individuals have access to science scale capital such as exposure to science hobbies (e.g. astronomy), out of school activities related to scale (e.g., museums, science centers), or adults with science or engineering careers.

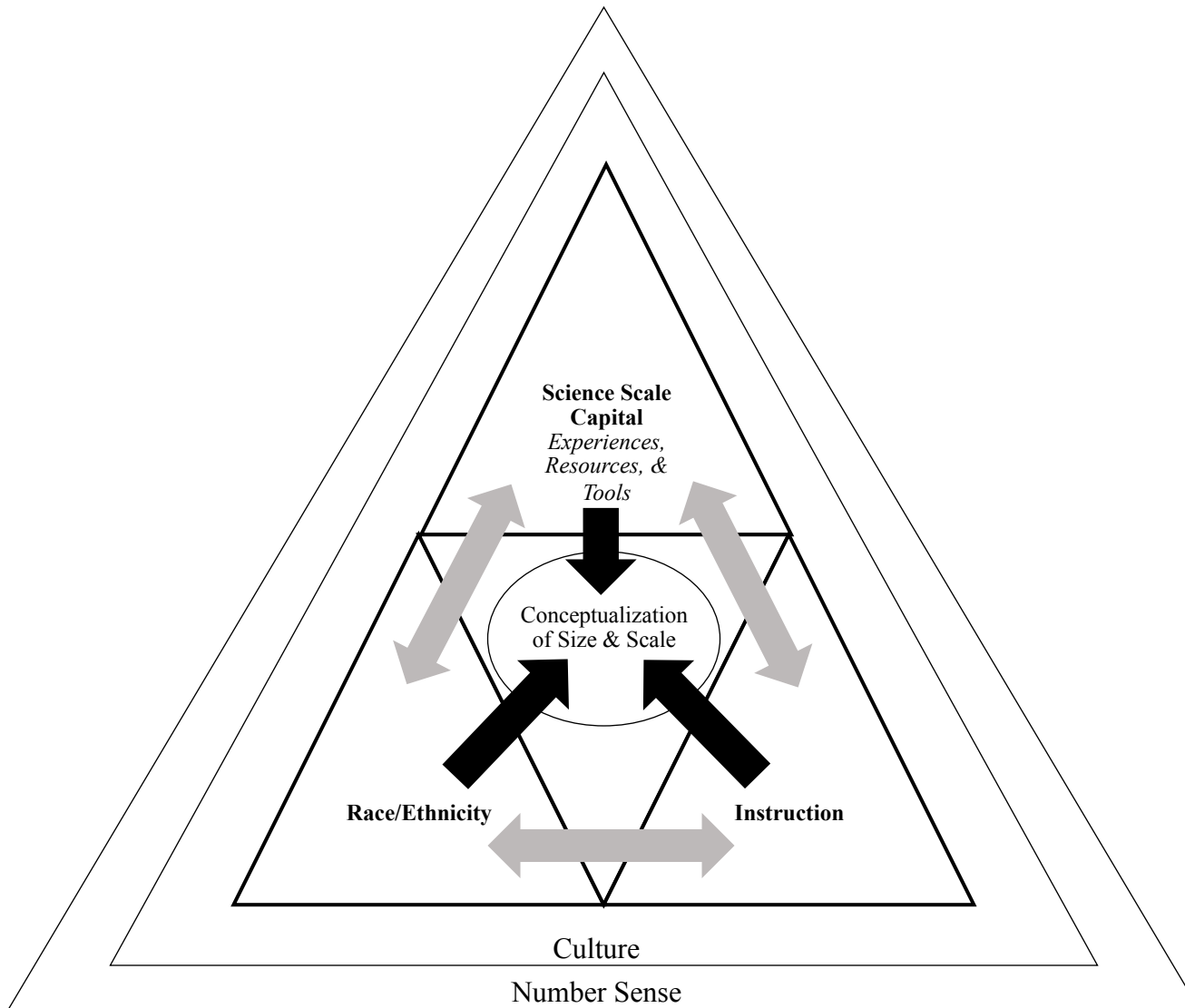


Figure 1. A model for factors that influence students' concepts of scale, which are situated within the context of both culture and innate sense of number.

Though size and scale comprise only one small fraction of all that is science, the theme becomes increasingly relevant as scalar models such as geologic time are necessary for understanding global climate changes or scalar models of the size and behaviors of nanoparticles are necessary for understanding new innovative medicines. In light of the need for a scientifically literate citizenry capable of driving both national and global policy, a basic understanding of size and scale as it relates to science is deeply needed. This research indicates that teaching about size and scale is within the grasp of educators and may be of benefit to explaining a wide range of scientific phenomena. Taking into account factors such as race/ethnicity, out of school experiences, and the use of explicit size and scale instruction, all of which are situated within the context of innate numerical abilities and cultural experiences, may enhance education for all students.