Computer-based Number Categorization as an Intervention for Computer-based Number Line Estimation

Ama Awotwi

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Computer-based Number Categorization as an Intervention for Computer-based Number Line Estimation

Ama Awotwi

The number line is a versatile tool. When used in estimation, it can serve as a visual representation of number. This study evaluated the relationship between sorting numbers by magnitude and number line estimation performance. Fifty-eight participants in Grades 1, 2, and 3 estimated values on a 0-100 number line over four sessions. During two intervention sessions they sorted numbers into 5 categories either linearly or nonlinearly before they estimated the same target values. The linear group's number line estimates had less error than the nonlinear group's estimates at posttest. In particular, the participants who started with low numeracy scores in the linear group outperformed their counterparts in the nonlinear group on the number line estimation task. Computer-based number categorization supports computer-based number line estimation skills when numbers are categorized linearly. This finding extends the *representational mapping hypothesis* to computer-based scaffolds.

Keywords: number line estimation, number categorization, magnitude, representational mapping hypothesis, computer-based

Table of Contents

List of Charts, Graphs, Illustrations

Acknowledgements

This dissertation could not have happened without the help and support of a community of people to whom I will forever be grateful. A special thank you to my advisor, Dr. Herb P Ginsburg without whom I would have never embarked on this journey or believed that I could see it through to the end. Your support was always exactly what I needed. I am thankful to Dr. Elizabeth Tipton for always making time to meet with me and for thinking outside the box to help me analyze my data. To Dr. Ryan Baker thank you for your insightful questions and feedback. Thank you to my committee members: Dr. Erica Walker and, Dr. Janet Metcalfe for making the dissertation defense a wonderful learning experience. Thank you for pushing my thinking further than I thought it could go.

Thank you to Dr. M and the GTD crew. Without you all and the many pomodoros I would not have gotten this done. Thank you for teaching me that writing can be a comfortable process. And also thank you for reminding me that life takes courage.

Thank you to Dr. Cassie Freeman, your support came in every possible form. Thanks for the many conversations and for always pushing me to the next level. I hope you know what an invaluable resource you have been - words cannot capture it.

I would like to express my gratitude to several members of the Teachers College community. To Katie Embree for your support over the years at some of my toughest moments. Thank you for always having an open door. To Laurie Behrman and the Human Development administrative team, thank you for helping me plan and execute all of my presentations. To the facilities staff and to public safety, thank you for taking care of my space, keeping me safe on the nights I worked late, and always greeting me each day with a smile. You help make TC more than a welcoming environment; you make it feel like home.

I would like to thank the MathemAntics team: Kara Carpenter, Samantha Creighan, Ben Friedman, Azadeh Jamalian, Rachael Labrecque, Dana Pagar, Esther Yoon for leading the way and showing me how this process is done. Also thank you for the discussions about software and experimental design. And thanks to the present members of Dr. Ginsburg's Lab, especially Colleen Uscianowski and Mia Almeda, for listening to my presentation and giving helpful feedback.

Thanks to my family. I would not be here without you.

To Alison Elmer, you help me remember what this work is really about. I appreciate knowing that with you I can always ask another question. To Sean Brooks thank you for reminding me to take breaks and see the world around me. To Jenna Tutjer thank you for the endless encouragement and for problem solving with me. To Jessamyn Martin, your friendship and support help me stay on track. Thank you for believing in me even more than I believe in myself. Thank you to my friends Abbie Wreibel, Kate Keating, Lissa Pressfield, Loriana Berman, and Mithu Tharayil who have been tireless cheerleaders throughout my schooling. Your willingness to read a draft or listen to me talk through my ideas is appreciated. Our friendship means the world to me.

Chapter 1

Introduction

Mathematics is all around us. Numbers, but one segment of mathematics, are omnipresent in our culture. Whether it is locating a building, grocery shopping, taking public transportation, or visiting the doctor, the use of numbers is inevitable. Further, it is not enough to only know how to count; numbers must have meaning beyond rote counting. Children come to school with informal numerical knowledge and need to learn formal mathematics, "a written, codified body of material conventionally defined and agreed upon" (Ginsburg, 1997, p. 23). How is numerical knowledge learned? What tools do children use to translate informal numerical knowledge into understanding about larger exact quantities?

Core Knowledge

A variety of theories exist to explain numerical development and how number is understood. Feigenson, Dehaene, and Spelke (2004) propose the presence of two core systems that comprise the foundation of numerical understanding. The two systems behave slightly differently for infants and adults. Core system 1 addresses approximate representations of numerical magnitudes. Core system 2 regards the precise representations of distinct individual numbers.

Core system 1.

Large numbers can be discriminated without counting and such discriminations "are ratio-dependent and robust across modalities" (Feigenson *et al*., 2004, p. 309). The ability for babies to discriminate between large numbers of items depends on the quantitative relationship between the numbers (i.e. the ratio) and does not depend on the type of item: for example they could be discrete objects or tones. Research controlling for continuous variables, such as perimeter, area, and density, shows that infants are able to distinguish between large numbers of things (e.g. 8 and 16 dots) with ratios like 1:2 (Xu & Spelke, 2000). Youths and adults are also able to judge large numerosities without counting regardless of modality (e.g. auditory or visual) (Barth, Kanwisher, Spelke, 2003) up to the ratio 7:8.

Core system 2.

Small numbers of individual objects can be tracked by preverbal children. Infants judge the absolute number of items and can distinguish between 1 versus 2 and 2 versus 3. In studies in which infants see objects hidden, babies search for discrete objects up to and including the number 3. In adulthood, as with infants, people are especially skilled at enumerating 1 to 4 items, which is known as subitizing: enumerating when there are fewer than four items (Trick & Pylyshyn, 1994). Though some adults can subitize up to 7 items, adults' are able to recognize 1 to 4 items rapidly and with an accuracy that greatly decreases for more than 4 items.

Core systems 1 and 2 are independent systems, yet Spelke and Kinzler (2007) cite three overlapping properties regarding number representations that the systems share: number representations are 1) "imprecise, and their imprecision grows linearly with increasing cardinal value" (p. 90); 2) abstract; 3) comparable and can be combined. Feigenson, Dehaene, and Spelke (2004) cite a limitation of their theory as being the lack of reference to the development of exact large numbers, or "fractions, square roots, or negative numbers" (p. 307). An additional limitation of this theory is that it is comprised of two separate systems that differ in adults and children. Ideally a theory would be integrated and have continuity across the developmental spectrum.

"Darwinian Competition" Theory

Geary (2006) frames the process of mathematical development as analogous to the processes of evolution. Children's mathematical competence and general cognitive development can be understood through consideration of variability, competition, and selection (Geary, 2006). Goal activation and activation of goal-related information dictates where attention is directed within the problem-solving space. Unlike the Core Knowledge Theory, this theory allows for many competing problem-solving strategies. "Multiple facts, concepts, and procedures are simultaneously activated" (p. 800), and from this the need for inhibiting less useful information arises. Not all of this process is explicit. At the neuronal level competition occurs for which neuronal groups will be activated and this leads to variation in the expression of strategies.

A key part of the mathematical development process is consistent selection of advantageous (i.e. successfully goal-achieving) strategies. Geary states that associative memory, memory in which an association is formed between two or more things (e.g. strategies), is one integral component of strategy selection and it leads to "greater variation in the number of processes that can be used to achieve a goal, thus greater competition among the processes" (p 801).

The Darwinian Competition theory can be seen in the development of basic addition skills. Early addition strategies involve counting all of the items in each addend. A student may eventually adapt to "counting on": starting from largest addend and counting the smallest addend by ones. In the final stages of the evolution of basic addition a student can retrieve sums from memory without counting at all. According to Geary (2006), "direct retrieval eventually gains a selective advantage over execution of counting procedures," (p. 801) most likely due to speed of retrieval and lower working memory demands. The speed of retrieval and accuracy lead to memory being a more reliable process than counting strategies – "the goal is achieved before execution of alternative processes" (p. 801). Goal achievement may reinforce the selected strategy (i.e. retrieval) and inhibit alternative processes.

Integrated Theory

While Geary's evolutionary theory can include all types of numbers (e.g. fractions, negatives), it does not provide specifics for early number development other than to delineate it as comprising biologically primary competencies. (For examples see Geary 2006.) Siegler, Thompson, and Schneider (2011) propose a theory of numerical development that "emphasizes a key developmental continuity across all types of real numbers" (p. 274). Their *integrated theory of numerical development*, which Siegler (2016) later expounds on and expands, outlines numerical development from non-symbolic to rational numbers. The theory has four overlapping types of changes that occur over development and can be summarized in six main claims, which will be stated here. The overarching theme of the theory is that numerical development is the "generation of increasingly precise magnitude representations for an increasingly broad range of numbers" (Siegler, 2016, p. 342).

According to the theory the four overlapping outcomes of the developmental process are: "(1) generating increasingly precise representations of numbers expressed non-symbolically; (2) connecting symbolic to non-symbolic representations of the magnitudes of small whole numbers; (3) extending the range of whole numbers whose magnitudes can be accurately represented; (4) progressing beyond whole numbers to accurately represent the magnitudes of an increasing range of rational numbers" (p. 342). The integrated theory can be summarized as the following six points:

1. All rational numbers are represented on a mental number line

2. Magnitude representations go from compressed to linear within certain number ranges with smaller numbers becoming linear earlier than larger ones
3. The knowledge that real number properties do not apply to all numbers, but that all numbers possess magnitude is key to rational number development
4. Association and analogy are two of the many processes that are essential to numerical magnitude development
5. Magnitude knowledge is at the foundation of numerical development; thus magnitude knowledge is "correlated with and causally related to other aspects of mathematics" (p. 343)
6. Interventions designed to improve magnitude knowledge will have positive effects on other aspects of mathematics (Siegler, 2016).

The integrated theory encapsulates the previously mentioned theories of numerical development and improves upon them by incorporating all rational numbers. Further, the theory captures numerical development from birth to adulthood. Siegler (2016) acknowledges that numerous processes affect numerical development and acknowledges the role of other mathematical domains in numerical magnitude development, such as spatial reasoning. The intersection between numerical development and spatial reasoning will be addressed in later sections.

Numerical Development

Despite the differences between the theories of numerical development, at the core they all share common mathematical tenets regarding counting and numerical magnitude. Common principles of enumeration exist across all theories of numerical development. Gelman and Gallistel (1978) outline five counting principles that bring clarity and precision to discussions regarding enumeration. The five principles are: one-to-one, stable order, cardinal, abstraction, and order irrelevance. Each is described here.

In the *one-to-one principle* each item to be counted must correspond to one and only one tag (e.g. count word). In order for this to occur enumerators must simultaneously track the items to be counted along with those that have already been tagged. Furthermore, the tagging system will only be effective if the counter has a set of unique tags from which to draw upon.

Having a unique set of tags is not sufficient when counting; the order of the tags must stay constant. For example, if a young child consistently uses the words 'one', 'two', 'three' to count three objects, yet changes the order of the counting words, the final count word will be different. The *stable order principle*, which states that the order of tags stays constant, is essential for *cardinality*.

When counting the last tag given to an item is the *cardinal number* and is used to represent the entire set: *the cardinal principle*. This principle is contingent upon the one-to-one and stable order principles. To successfully apply the cardinal principle enumerators must track the final tag and be able to respond with such when asked the question 'How many?'

The *abstraction principle* defines to what the preceding principles may be applied or, in short, what can be counted. Any set of discrete objects can be counted and the items may be counted in any order, which is the *order irrelevance principle*. Gelman and Gallistel (1978) state "the same cardinal number results regardless of order of enumeration" (p. 82).

To move beyond rote counting enumerators must connect counting words to magnitude. Children must know that "each word's position…relates directly to its meaning – the farther along a word occurs in the list the greater numerosity it refers to" (Wynn 1992, p. 220). This is the concept of *ordinality*: “successive number words represent successively larger quantities” (Geary, 2006, p. 786).

Furthermore, the connection must be made between numbers, symbolic numbers (e.g. Arabic numerals), and magnitude. The intersection of numerals and magnitude is the mental number line.

Mental Number Line

Moyer and Landauer (1967) demonstrated that when comparing two single digit numbers, participants took longer to make judgments when the numbers were closer together in magnitude. Additionally, participants made more errors when the numbers being compared were closer together in magnitude. Moyer and Landauer (1967) state that “the results strongly suggest that the process used in judgements [*sic*] of differences in magnitude between numerals is the same as, or analogous to, the process involved in judgements [*sic*] of inequality for physical continua” (p. 1520), such as the lengths of lines. In other words, mentally comparing number is like comparing actual physical distances, suggesting that number is situated on a mental number line and different numbers correspond to different distances.

In an experiment similar to Moyer and Landauer’s (1967), Restle (1970) had participants choose which was larger between the sum of two numbers (A + B) and a third number (C). Restle (1970) found that when A + B was far from C participants were faster and more accurate with making a decision. Further, when A was far from B speed and accuracy also increased. When the difference between A+B and C was small and required more precision, time to answer increased, as did the number of errors. Restle hypothesized a model of participants using “a number line, an analog system having distinctive markers,” (p. 277) to make magnitude judgments.

The patterns that Moyer and Landauer (1967) and Restle (1970) observed later came to be known as the distance effect and the magnitude effect. The distance effect specifies that it is easier and quicker to judge comparisons of numbers that are far apart as opposed to closer together (Dehaene, 2011). For example, it is easier to judge which is larger between 10 and 60 than to judge between 63 and 65. The magnitude effect is that for pairs of numbers that are equidistant the smaller pair will be easier to discriminate (Dehaene, 2011). So if given 3 and 5 or 63 and 65, people will take less time and be less prone to errors when judging the pair 3 and 5.

Another well-documented effect that demonstrates the connection between number and spatial orientation and points to the existence of a mental number line is the *Spatial-Numerical Association of Response Codes* (SNARC) effect. Dehaene, Dupoux and Mehler's (1990) study led to the discovery of the SNARC effect. In the study participants compared numbers and used either their left or right hand to indicate which digit was larger. Participants responded faster and more accurately when the larger number corresponded to the right and the smaller number was on the left side. The SNARC effect is seen even when the task is a parity task: judgment of odd or even and magnitude is irrelevant (Dehaene, Bossini, Giraux, 1993). The SNARC effect suggests that people's mental orientation of number is with larger numbers on the right as with a number line. However, it is important to note the SNARC effect as described is found in cultures with left-to-right reading and writing orientation (see Ito & Hatta; Shaki & Fischer, 2008; Zebian, 2005 for examples of other orientations in other cultures).

The SNARC effect as evidence of the existence of a mental number line has detractors. Santens and Geevers (2008) conducted a study that calls into question the argument that direct mapping occurs between numbers and direction. The study had participants respond unimanually and unidirectionally (left or right) to say if a number was close or far from a predefined standard number. Participants were faster when close corresponded with smaller numbers and far corresponded with larger numbers regardless of direction. The authors present this as evidence that the spatial-numerical association is best explained by “an intermediate categorization of numbers as relatively small (- polarity) or large (+ polarity)” (p. 269). It is the intermediate categorizations that are linked to the response codes (i.e. close/far).

Santens and Geevers (2008) hypothesize that the SNARC effect is not evidence of the existence of a mental number line and rather an intermediate step that exists between the spatial-numerical association and the response codes. However, the evidence in support of a mental number line has been observed on a behavioral level as well as is well documented at the neuropsychological level (see Hubbard, Piazza, Pinel & Dehaene, 2005 and Umiltà, Priftis, & Zorzi, 2009 for reviews). Thus, this paper operates from the premise that the mental number line does exist and is a useful construct.

While the SNARC effect posits the mental number line, it does not explain the origin of the mental number line. Number lines are directly taught in formal schooling as a tool for tracking and solving mathematical operations. In the words of Krasa and Shunkwiler (2009) “Number lines can also illustrate the conceptual underpinnings of nearly all elementary number skills… the number line provides a schematic image or mental template that children can rely on and abstract from” (p. 28). Additionally number lines are present in school indirectly, like in charts for tracking the days of the school year. Number lines are also visible in objects whose primary feature is not to teach number or mathematics (e.g. number keys on a computer keyboard). It is possible that the mental number line exists because people are inculcated with number lines through school and their daily environment.

Number Line Estimation Task

The number line estimation task is the embodiment of the mental number line. The number-to-position (NP) task is done on an open line that typically has its endpoints marked with 0 and 10, 0 and 100, or 0 and 1000. The task requires estimators to place target values on a line, usually one estimate at a time. As Siegler and Ramani (2011) explain, "the number line estimation task involves asking [people] to translate between numerical and spatial representations" (p. 345). It is not surprising that number line estimation is considered both a numerical and spatial task since theorists propose that a close relationship exists between number and space (Mix & Cheng, 2012) and that relationship is evident during the number line estimation activity.

Number line estimation incorporates elements of number, such as magnitude and symbolic number (i.e. Arabic numerals). Additionally, elements that are integral to comparing numbers, such as proportion and scale, are also present in the number line estimation task. Siegler and Ramani (2008) write that one "advantage of this task is that it transparently reflects the ratio characteristics of the formal number system" (p. 665). For example, 30 is half of 60 and is located half as far from zero on the number line. Many mathematical elements are incorporated into the number line estimation task, yet the task's design is simple and does not require outside knowledge (e.g. units), only the mental representations that are at the foundation of number.

Mental representations.

What are the underlying mental representations of number that are captured by the number line estimation task? Theorists do not agree on one mental representation. One view is that there are multiple different representations at play during number line estimation and the representation used depends on context and on development (Siegler & Opfer, 2003). This multiple representation view proposes a developmental shift from estimates following a logarithmic curve to fitting a linear function (Siegler & Booth, 2004). The logarithmic mental representation is prevalent for a number range (e.g. 0 – 100) when children view smaller numbers as being further apart than actual and perceive larger numbers as being closer together. For example, the estimated location of 15 is near where 60 is actually located (Siegler & Ramani, 2008). Gradually, with experience, the mental representation shifts to a linear representation (Ashcraft & Moore, 2012; Siegler & Booth, 2004) in which a one-to-one relationship exists between the number estimated and the target number (i.e. 15 is estimated near its actual location).

Another established perspective on the mental representation for number is grounded in the idea that number line estimation is a proportion-judgment task (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011). The model is based on the premise that the number line estimation task requires consideration of part-whole relationships: the 'part' is the target number and the 'whole' is the range. Barth and Paladino's (2011) proportion-judgment model, which extends the cyclical power model (Hollands & Dyre, 2000), proposes a power function to best describe estimates graphed against actual magnitudes. A power function captures a key feature of number line estimation data: the "systematic patterns of over- and underestimation" (Barth & Paladino, 2011, p. 126). Similar to the logarithmic-linear model, the proportion-judgment model has two parameters: β, the exponent which determines the shape of the power function that relates actual magnitudes and estimates, and W, which represents the subjective scale children use and "accounts for young children's lack of knowledge of the magnitude of 100" (Barth & Paladino, 2011, p. 130). The model accounts for use of anchors or reference points (e.g. the midpoint) and misjudgment of the whole range (Barth & Paladino, 2011). While the model can account for subjective judgments of the scale of the whole range, Barth and Paladino (2011) acknowledge that the model does not incorporate all of the variance that must be present in the estimates. In the authors' words, "children ignorant of the magnitude of '100' are unlikely to judge the presented numerals relative to a single stable whole magnitude across trials" (Barth & Paladino, 2011, p. 130). Further, it may be possible that children do not attend to the highest endpoint and thus do not use it when estimating. In spite of the model's shortcomings, some theorists still consider the proportion-judgment model the best model for the underlying mental representation of number exhibited in number line estimation tasks (see Cohen & Blanc-Goldhammer, 2011; Slusser, Santiago, & Barth, 2013; Sullivan, Juhasz, Slattery, & Barth, 2011)

Some theorists are proponents of a segmented linear model as the mental representation of number underpinning number line estimation (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008; Moeller, Pixner, Kaufmann, & Nuerk, 2009; Moeller & Nuerk, 2011). The segmented linear model proposes using at least a two-phase linear model, two different regression line segments, to account for children's understanding of number. One line segment fits familiar numbers, while another segment with a lesser slope fits the unfamiliar numbers. The "change point", the point where the two line segments break, is considered the point at which children began to discriminate less between numbers (Ebersbach et al., 2008). (Moeller et al. (2009) propose a fixed change point at the number 10.) Thus, Ebersbach et al., (2008) propose that a segmented linear model, in addition to being a simpler model, provides more useful information than the dominant logarithmic-linear model.

Regardless of the underlying structure, theorists agree that the locations of larger numbers are more difficult to estimate than smaller numbers. Furthermore, regardless of the mathematical function that is used to describe the shape of the mental representation in its early stages, theorists are in agreement that a linear function or a linear-looking function is the ideal mental representation. Finally theorists concur that the representation of number improves with numerical magnitude understanding. The number line estimation task measures the degree of linearity in this mental representation, which is one reason it is such a valuable activity.

Number line estimation and achievement

In addition to being a measure of mental number representations, number line estimation as a numerical magnitude task is important for its connections to mathematics success. Cowan & Powell (2014) found that number line estimation accuracy is related to a range of arithmetic skills (i.e. calculation fluency, written arithmetic, word problems). Beyond the relationship between number line estimation accuracy and arithmetic, number line estimation performance is also related to mathematics achievement more broadly.

Various studies (Ashcraft & Moore, 2012; Booth & Siegler, 2006; Siegler & Booth, 2004; Sasanguie, De Smedt, Defever, Reynvoet, 2012) have documented the relationship between number line estimation and standardized mathematics tests. Sasanguie et al., (2012) found an association between number line estimation and curriculum-based standardized mathematics achievement test in grades K, 1, 2, and 6. The authors noted that the more linear estimates were associated with higher scores on the mathematics achievement test (Sasanguie et al., 2012).

The associations between number line estimation accuracy and mathematics success are not only evident in the short term (see Krasa & Shunkwiler, 2009). Functional numeracy measures are assessments that tend to be more focused than achievement tests in that they assess "mathematical competencies that influence economic opportunity and other real-world outcomes" (Geary, Hoard, Nugent, & Bailey, 2013). Early number system knowledge, which includes tasks like number line estimation, predicts performance on functional numeracy measures as much as six years later (Geary, et al, 2013). Geary et al. (2013) found that these results occur even after controlling for other important factors, such as intelligence, working memory, general mathematics achievement, and demographic statistics. The connection between early number system knowledge and later scores on functional numeracy measures suggests that number line estimation accuracy is essential beyond success in formal mathematics within school. The skills captured through number line estimation are an indicator for later life success.

Number line estimation skills and strategies.

What skills are needed for number line estimation? Sullivan, Juhasz, Slattery, and Barth (2011) identify that the "…[task involves] at least three number-related components: observers' understanding of symbolic number systems, mental representations of numerical quantity, and strategies for mapping numerical information onto space" (p. 557). Let us examine these three components further. Estimators must be able to read symbolic number (e.g. Arabic numerals) and map magnitude onto the numbers they read. For example, if shown the number 63 an estimator would need to be able to read the numeral and have a sense of how much 63 is. The estimator's "mental representation of numerical quantity" is the skill that allows her to draw meaningful relationships between different numerals, such as knowing that 63 is more than 50 and less than 100. Finally, Krasa and Shunkwiler (2009) suggest an association exists between spatial skill and number-line placement accuracy. Specific strategies for mapping number onto space will be discussed in detail in the following section.

Similar to the range of skills encompassed in number line estimation, a range of strategies exists for performing number line estimation. Some strategies are more effective than others, people's degree of cognizance about strategy usage may vary, and no evidence suggests that people exclusively use one strategy to estimate a target value. Siegel, Goldsmith, and Madson (1982) conducted an estimation study in which they identify the strategies used by estimators. While the authors' study involved estimating length, height, and discrete quantity (numerosity), some of the ten strategies they identify still apply here. *Benchmark comparison*, also called *anchor* (Ebersbach et al., 2008; Siegler, 2016), *landmark* (Siegler & Opfer, 2003), *reference point* (Cohen & Blanc-Goldhammer, 2011), or simply *benchmark* (Ebersbach et al., 2008), is when comparison to another distance is used to estimate (Siegel et al, 1982). For example, a child may know the midpoint of a range and use that information to estimate numbers that are near the midpoint. Endpoints may be used as reference points. The use of landmarks demonstrates use of the proportional relationships referenced in the proportion-judgment model. However, it is not clear that proportions are used exclusively. For example, to estimate the number 53 a child may use the midpoint of the 0-100 range to orient herself, then may count up 3 spaces to get to 53. *Decomposition/Recomposition* is another strategy and involves breaking the target value into parts and then recombining parts (Siegel et al., 1982) in order to place the estimate. For example, to estimate the number 40, a child may estimate ten four times.

The distinctions between number line estimation strategies need further study. However, it is clear that the use of certain strategies results in more accurate estimates than others. Crites (1992), using the same strategy categories as Siegel et al. (1982), interviewed estimators and found that more highly skilled estimators used benchmarks and decomposition/recomposition, whereas less skilled estimators tended to not verbalize a strategy, to guess, or to inaccurately decompose. Ashcraft and Moore's (2012) study of participants in Grades 1 through 5 and college-aged adults demonstrates a progression of strategies used by estimators. By analyzing variability in errors it appears participants in Grade 1 estimated from the origin, in an "origin-up"

(p. 266) strategy, whereas participants in Grade 2 estimated from both endpoints. By Grade 3 evidence suggests participants are using the midpoint as well as the endpoints to estimate, a trend that continues and strengthens into adulthood. The developmental trajectory of strategies from the origin-up strategy to using the midpoint would suggest that perhaps Barth and Paladino's (2011) proportional judgment model would not apply to all developmental groups of estimators. Beyond using the midpoint, data indicate that people may divide the number line into quarters (see Siegler & Opfer, 2003). Siegler & Opfer (2003) comment that, "… the number of landmarks may vary with task characteristics, but… relying on subjective landmarks is one effective strategy for generating linearly increasing estimates" (p. 242). Cohen & Blanc-Goldhammer (2011) also noted that the standard deviation of estimates was lower near reference points.

Building number line estimation skills and strategies.

Recognizing the importance of number line estimation, a few studies have sought to improve number line estimation skills or the underlying magnitude knowledge that number line estimation measures. Researchers have attempted to build number line estimation accuracy primarily through feedback and games. Below is a brief summary of several studies that focus on building number line estimation skills.

Opfer and Siegler (2007) provided feedback to second grade participants at several points along the number line including at the point of greatest discrepancy between logarithmic and linear curves, 150, on a 0-1000 number line. The authors hypothesized that feedback at the greatest discrepancy point would lead to the greatest change because they supported the logarithmic-linear model as the mental representation underlying numerical development. Participants who received feedback were able to shift their representation from logarithmic to linear, with those who received feedback at 150, the point of greatest discrepancy on the 0-1000 range, making the shift faster than those who received feedback on other numbers. From their study Opfer and Siegler (2007) conclude that "feedback on numerical magnitudes [is] a potent source of change" (p. 190).

Siegler & Ramani (2008) attempted to improve low-performing children's number line estimation skills by introducing a linear numerical board game that children played in 15-minute sessions. The alternative to playing the linear numerical board game was to play a linear game with colors instead of numbers. The children who played on the numerical game board improved their number line estimation linearity and accuracy. The best fitting linear function accounted for 52% of variance at pretest and 96% of variance at posttest. *Percent absolute error* (PAE) (see Siegler & Ramani, 2008), decreased from 28% to 20%, $t(17) = 2.43$, $p < .05$, $d = .71$, demonstrating a significant improvement in accuracy. The children who played the color board game did not show improvements on linearity (R^2 was 73% at pretest and 36% at posttest), nor did number line estimation accuracy improve (PAE was 27% vs. 28%). Thus, playing the color board game did not impact children's understanding of numerical magnitudes, whereas playing the linear numerical game significantly improved knowledge of numerical magnitudes.

In 2009, Siegler and Ramani again tested linear numerical board games. However this time they contrasted linear numerical board games versus circular numerical board games and a group that did numerical activities. The number line estimates for the children in the linear board game condition were significantly more linear and more accurate. The best fitting linear function accounted for 22% of variance at pretest and 94% of variance at posttest. Accuracy was again measured as PAE and for the linear board game participants absolute error decreased from pretest (29%) to posttest (21%) ($t(29) = 4.85$, $p < 0.001$, $d = 1.01$). The group that played the circular board game had a small change in linearity (11% vs 26%) and a small but significant decrease (improvement) in accuracy (29% to 26%, $t(28) = 2.14$, $p < 0.05$, $d = 0.43$). It is in this study that Siegler and Ramani (2009) propose the representational mapping hypothesis: "The greater the transparency of the mapping between physical materials and desired internal representations, the greater the learning of the desired internal representation" (p. 547). In this instance the desired internal representation was a linear number line, so a linear board game helped to improve learning that internal representation.

Creighan (2014) investigated how MathemAntics Number Line (MANL), a digital number line activity, promotes number sense in Grade 2 children and particularly about the use of a flexible response range, the user-defined range (UDR). Suppose that a child sets the manipulatable range very wide, for example 20 numbers above and 20 numbers below the estimate number. If the target number is 41, and the child estimates that the number is 22, the estimate will be considered correct because it is within the ±20 answer range. But in the same situation, if the child sets the range to ± 2 of the target number, and estimates that it is 44, she will be wrong because 41 lies outside the range. A wide range is extremely tolerant of error and wary of risk; a short range is risky and intolerant of error. This approach differs from the traditional paper-pencil number line estimation task that asks participants to locate a point on the line.

While no effects were found on the number sense measures used in Creighan's (2014) study, children's performance on MANL improved over time. Further, Creighan (2014) found that prior number sense ability affects how children engage with MANL. Children with low number sense ability performed better with the fixed response range (FR) and children with high number sense ability performed better with the UDR. The children in the UDR condition, regardless of having low or high number sense ability, used better expressed strategies than the children in the FR condition. These findings suggest that MANL UDR condition helped some children improve their number line estimation abilities and helped build expressive strategies.

One study (Laski & Siegler, 2007) used a combination of a game-like activity and feedback to improve number line estimation skills. In this study participants categorized numbers as *Really Small*, *Small*, *Medium*, *Big*, and *Really Big* and some participants received feedback on the accuracy of their categorizations. The authors found that number categorization followed a logarithmic to linear developmental progression similar to that of number line estimation. The authors also found that number categorization improved regardless of the provision of feedback. Finally, although there was no training on number line estimation, the authors also found that after number categorization with feedback the number line estimation linearity also improved. Laski & Siegler (2007) attribute the change in number line estimation linearity to the "divide and conquer" approach. Number categorization helps divide the number line into five equal size categories decreasing the cognitive demand of the full, undivided number line. Further, I propose number categorization may support the "generation of increasingly precise magnitude representations for an increasingly broad range of numbers" (Siegler, 2016, p. 342), the crux of the *integrated theory of numerical development*. Number categorization helps build associations between symbolic number and the objective numerical categories into which the numbers are sorted. As stated above, one of the six main points of the *integrated theory of numerical development* is that the building of associations is key to numerical understanding. Additionally, number categorization encourages the use of efficient strategies to estimate on the number line by demonstrating that the number line can be decomposed into equal linear categories. It moves estimators beyond "origin-up" (Ashcraft and Moore, 2012) or "endpoint-down" strategies toward more advanced strategies (e.g. benchmark comparison, decomposition/recomposition).

Research Questions

The five studies summarized above suggest it is possible to improve number line estimation skills. This study extends the previous studies to answer several questions: *How do children perform on a computer-based number categorization task? How do children perform on a computer-based number line estimation task? How does number line estimation performance change over time? How does number categorization affect number line estimation? How do individual differences affect number categorization and number line estimation?* The following sections will briefly describe the study and address each question.

The study

The study incorporates a computerized number categorization task into a number line estimation task. One group of children, the *Linear* condition, saw 5 boxes horizontally arranged in a row. On each box was written one of the following words in the following order: *Really Small*, *Small*, *Medium*, *Big*, and *Really Big*. A second group of children, the *Non-Linear* condition, saw the same 5 boxes except arranged in a configuration that is a 2 by 2 array with the *Medium* box in the middle. After categorizing a number children estimate the location of that number on the number line.

1. How do children perform on a computer-based number categorization task? 2. How do children perform on a computer-based number line estimation task?

The study expands on Laski and Siegler (2007) by having a computerized number categorization activity and number line estimation task. Computers are able to support student learning and allow for exploration in a way that paper and analog tasks do not. Ginsburg, Jamalian, and Creighan, (2013) capture the benefits of software and scaffolds for building new strategies:

> "Software has unique affordances to encourage the use of strategies by highlighting the advantages of new strategies or limiting the resources needed to use another strategy…. Carefully designed scaffolds could also help a child in adopting a new strategy. Scaffolding may highlight the features of the problem that are relevant to a strategy, ease the difficulties involved in adaptation of a new strategy, and model the strategy for the child." (p. 97, 99)

The computerized number categorization may act as a scaffold for children's number line estimation by encouraging them to think about the entire number line from the really small numbers to the really big numbers. Additionally, the categories may promote more advanced strategies, such as decomposition/recomposition (Siegel et al., 1982) strategies.

The computer-based version of the activity also supports research into children's learning because of the rich backend information that the activity provides. The backend data, or log data, has record of where users click, the time it takes to complete each task, or trial, if they use help features, and a variety of other data. Within the log data are clues about estimation behavior. For example, if a child's initial click is near zero on all trials regardless of the target value, it could be surmised that the child is using an "origin-up" (Ashcraft and Moore, 2012) strategy. However, if the initial click location varies with the target value it could be hypothesized that multiple strategies are being employed. It should also be noted that the initial click location might not correspond to the first place the child uses to determine the estimate's location, so such data needs to be examined with caution.

Computer-based learning has a third potential benefit that may be particularly important for students of color and girls. Students are affected by perceived stereotypes in their learning environment (Steele, 1997). One prevalent stereotype is that mathematics is a subject for white males (Epstein, Mendick & Moreau, 2010; Oakes, 1990). While empirical research has not been collected to support this claim, a high quality computerized mathematics activity, which provides feedback solely based on input and not user demographics, may act as an unbiased learning tool that allows all students to thrive.

While a computer-based activity has the aforementioned affordances, it is expected that it will still measure performance at least as accurately as the analog version of the activity does. If the computer-based activity is measuring accurately then one would expect to see differences between the participants of different grades. On the number categorization task and on the number line estimation task the oldest participants should outperform the youngest participants mirroring the results seen in other studies (e.g. Laski & Siegler, 2007; Siegler & Opfer, 2003)

3. How does number line estimation performance change over time?

Studies of growth over time are essential to the study of cognitive development. Using multiple measurements over a brief period of time, similar to a microgenetic (Kuhn, 1995; Siegler & Crowley, 1991) approach, affords the ability to study mechanisms of growth and answer questions about which strategies are being used. Siegler (1998) describes the overlapping waves theory of strategy development: at any given moment people have multiple strategies available to them and strategy selection depends on the context of the problem-space. As learning occurs the use of some strategies decrease and while other, usually more efficient, strategies increase. An examination of the variability in number line estimates can provide clues into children's strategy use. For example, accuracy may be greatest and variability least for numbers near the origin, suggesting an "origin-up" (Ashcraft and Moore, 2012) strategy. As number line estimation skills improve, numbers near the endpoint and midpoint will have an increase in accuracy and decrease in variability suggesting an adaptation in strategy use, such as starting from the endpoint for higher numbers and using the midpoint as an anchor. For numbers in the final quintile (i.e. target values 80-100) if participants develop an "endpoint down" number line estimation strategy, then over the sessions accuracy should increase and the standard deviation should decrease.

4. How does number categorization affect number line estimation?

While analog number categorization helps number line estimation, it is not known whether categorization needs to be into linear categories or if categories can be non-linear and still beneficial to number line estimation. From Bächtold, Baumüller, and Brugger's (1998) study in which priming with a clock face led to a "reverse" SNARC effect as if numbers were mapped clockwise, and from Siegler & Ramani's (2009) study in which a linear board game was compared with a circular board game, one can conclude that the spatial orientation of how numbers are presented matters for some tasks. Does spatial orientation also affect number categorization? When describing the benefits of the "divide and conquer" approach Laski and Siegler (2007) write of participants "…using the categorizations to constrain number line estimates … by mapping the categories onto corresponding parts of the number line…" (p. 1741). From the representational mapping hypothesis (Siegler & Ramani, 2009) we know that "mapping" in the context of board games becomes easier or more difficult depending on the orientation of the representation that is being mapped. Applying the representational mapping hypothesis and extending it from board games to a computer-based categorization task, the study will evaluate whether linearity is a component of categorization that supports number line estimation skills by contrasting a linear with a non-linear categorization orientation. If categorizing numbers linearly supports number line estimation than the participants in the linear group should estimate more accurately on the number line than the participants in the nonlinear group. If the linearity of categorization does not matter than the two groups will estimate equally well.

5. How do individual differences affect number categorization and number line estimation?

Previous number line estimation studies (e.g. Ramani & Siegler, 2008) have shown the importance of pretest differences in predicting the response to interventions. Participants with low initial knowledge were still at the low end of performance at the study's close and vice versa. While other number line estimation studies (e.g. Siegler & Ramani, 2009) have shown that those who perform worse at the start make the larger gains after the intervention and catch up to their higher performing peers. And another number line estimation study (e.g. Creighan, 2014) showed mixed results in performance, with those participants who initially performed above the median for number knowledge continuing to perform well in particular contexts.

For the present study, initial individual differences in numeracy, as measured by a magnitude comparison task, and a numeral identification task may dictate student performance throughout the study. The participants who have lower numeracy and number identification skills at the start may make larger gains than the participants with higher numeracy and number identification skills at the study's outset. It is also possible that individual performance will depend on the context of the study: participants with low numeracy and number identification skills will perform well in certain conditions, while the participants on the higher end will perform well under other conditions.

Learning Study

Beyond the broader questions the study addresses, this is a project of software development and a "learning study" (see Ginsburg, Labrecque, Carpenter, & Pagar, 2015).

Learning studies, among other characteristics, "tend to include a small-to-medium sized sample and are highly focused on specific learning objectives or particular aspects of the software" (Ginsburg, Labrecque, Carpenter, Pagar, 2015). In game design, it is not enough to assume a spatial orientation for number categorization, a feature that is essential to the activity's learning experience, is beneficial. Different orientations must be tested so the one that most supports learning can be incorporated into the design, if indeed one orientation supports learning more than another. A learning study is needed to determine the effectiveness of the software and which orientation for the sorting categories is best. This particular learning study examines number categorization and the linearity of the boxes that serve as the categories for sorting in order to determine which boxes are effective scaffolds. If the linear boxes are more effective scaffolds than the nonlinear boxes, then the participants in the linear group will make gains above and beyond those of the nonlinear group on number categorization and number line estimation. If the nonlinear boxes are the better feature, then participants in the nonlinear group will make more significant gains on the number tasks. If the scaffolds perform equally well or detract from task success, the participants of both groups will make equal gains or will see a decrease in their performance respectively.

Chapter 2

Method

Participants

The study included 58 participants from Grades 1 through 3 at the start. As happens over the course of a study, attrition occurred. Table 1 presents the sample breakdown and mean age over the course of the four study phases.

Table 1. Participant sample and mean age

	Pretest	Session 1	Session 2	Posttest
n (boys, girls)	58 (29, 29)	48 (20, 28)	40 (16, 25)	29 (11, 18)
Mean age in months (SD)	7 years, 7 months (11.90 months)	7 years, 7 months (11.57 months)	7 years, 8 months (11.62 months)	7 years, 8 months (12.87 months)

The children were from a large urban school district. The city's schools are 48.6% female and are comprised of the following race/ethnicity demographics: 15.5% Asian, 27.1% Black, 40.5% Hispanic, 2.1% Other, and 14.8% White. However, most of the study's participants were Black and Latino. The two programs from which the participants were recruited were free summer programs in Harlem that serve low- to medium-socioeconomic status families within the neighboring community.

Procedure and Materials

Overview

Participants first met with a researcher one-on-one to complete a numeral identification task derived from mCLASS Math (©Wireless Generation) and a magnitude comparison task (Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013). These were included to establish a baseline of participants' basic number competency: numeral reading ability and magnitude comparison are key aspects of general number sense (Gersten, Jordan, & Flojo, 2005).

Participants also did 25 trials of a computer-based number line estimation task, MathemAntics Number Line Estimation (MANL). After the pretest, participants attended two sessions in small groups of 5 – 8 participants, during which time they completed the MathemAntics Sort and Estimate (MASE) number line task independently on a laptop. For the posttest, participants did the same MANL activity they completed during the pretest, to provide a comparison point. (See Table 2.) Each task is described in detail below.

Table 2. Tasks within each study phase.

	Pretest	Session 1	Session 2	Posttest
Numeral Identification	X			
Magnitude Comparison	X			
MANL	X			X
MASE		X	X	

Each participant used an Apple MacBook Air or MacBook Pro Laptop preloaded with the digital measures, MANL and MASE. Participants were given headphones to reduce the likelihood of distractions from fellow participants' audio. In previous pilot studies, in which young participants used laptops, some participants had difficulty controlling the cursor with the built in mouse control mechanism, the track pad. External mouses were offered to each participant to use as an alternative to the laptop's track pad. Each session participants elected to use either the mouse or the track pad depending on which they found to be more comfortable. At the end of each phase participants chose a sticker as a reward.

Tasks

Number identification and magnitude comparison.

Participants completed a number identification task during which they were asked to read aloud as many numerals as they could within 60 seconds (©Wireless Generation). After completing the number identification task, participants completed the numeracy screener, a two-minute paper/pencil task of symbolic (digits) and non-symbolic (dots) numerical magnitude comparisons (Nosworthy et. al., 2013).

MathemAntics Number Line Estimation.

Following the paper-and-pencil tasks, participants used computers to do the computer-based number line estimation task. MathemAntics Number Line Estimation (MANL) was one task in the estimation “environment” in a suite of computer-based mathematics activities, MathemAntics. MathemAntics was designed for students between 3 years old and Grade 3 (Ginsburg, Carpenter, & Labrecque, 2011). MANL allowed users to explore estimation using a horizontal number line. While the activity had many versions, the version participants used in this study had them determine the location of a given value on a horizontal line, a number-to-position task (see Figure 1). Participants indicated the estimate’s location with a horizontal range that was preset to 5 units, with 2.5 units on either side of a red, downward-pointing arrow. (See Figure 1.) MANL was not network- or Internet-dependent. The software had the capability to record user data in a log and captured information such as performance accuracy data, timestamps and precise estimate locations.

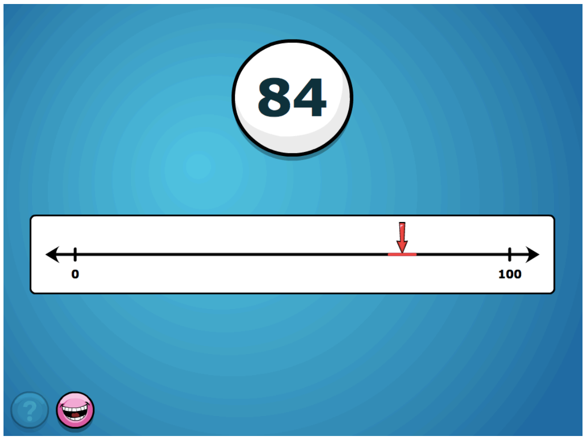


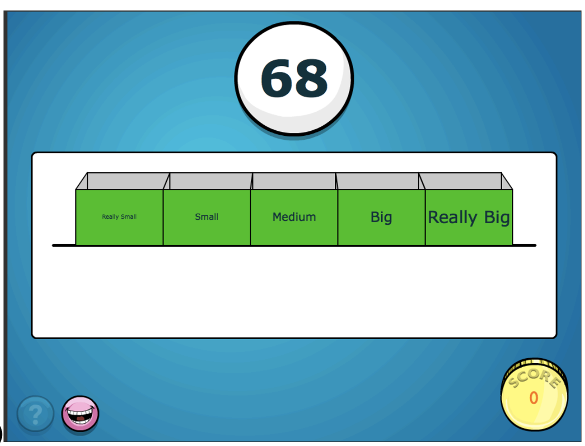
Figure 1. MathemAntics Number Line Activity.

Participants completed 25 trials with the values: 1, 7, 8, 13, 16, 20, 25, 27, 34, 35, 43, 44, 47, 56, 59, 61, 63, 66, 76, 77, 81, 84, 87, 92, 96 presented in random order. The numbers were selected for balance across the number line: there were five numbers for each range of 20, or quintile. On each trial one of the above written numbers was shown 4.5 cm above the 15 cm long number line and participants had one attempt to estimate the number's location on the line. Twenty-five questions was a typical number in comparison to previous estimation studies, which had 22 test trials (Laski & Siegler, 2007) and 26 trials (Booth & Siegler, 2006). Unlike the cited studies, this study did not oversample numbers in the first few decades because one goal was to have a balanced number of trials across the number line in order to have a more accurate measure of performance across the number line

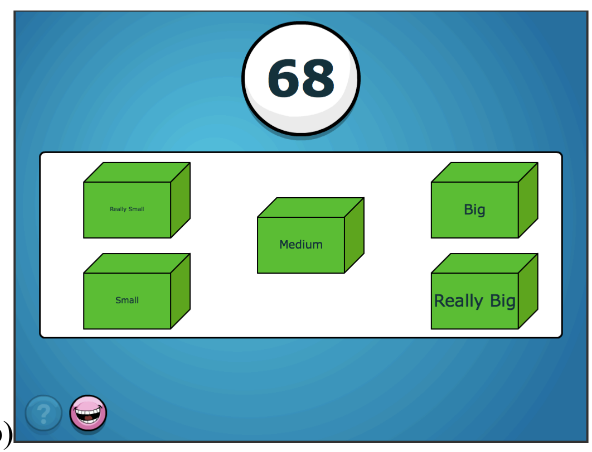
MathemAntics Sort and Estimate.

MathemAntics Sort and Estimate (MASE) was a three-part activity. First, participants watched a demonstration video that oriented them to the activity's range, the boxes' ranges, and gave instructions on the activity's controls.

a)



b)



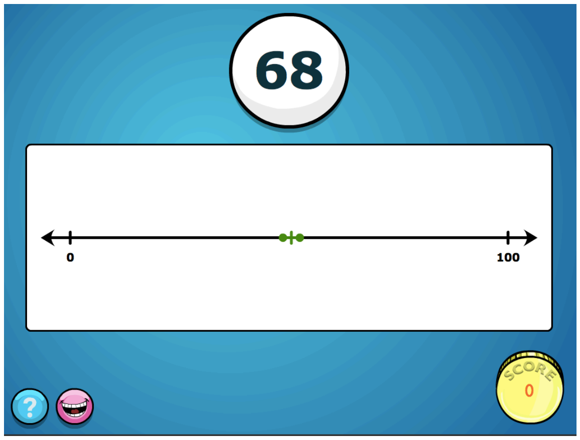
c) 

Figure 2. a) *Sort: Linear* activity of MASE; b) *Sort: Nonlinear* activity of MASE; c) *Estimate* activity of MASE

Second, after the demonstration video, MASE then had a magnitude categorization task, *Sort*. During *Sort*, participants were shown a number and were asked to indicate in which of 5 boxes a number belongs. The 5 boxes were labeled *Really Small*, *Small*, *Medium*, *Big*, and *Really Big*. The narrator gave the instructions: "Here are boxes labeled "really small", "small", "medium", "big", and "really big". Click the box where n belongs?" The number could only be placed in one box. Participants saw one of two versions of *Sort*: *Linear* or *Nonlinear*. In *Sort: Linear* the 5 categorical boxes (i.e. *Really Small*, *Small*, *Medium*, *Big*, and *Really Big*) spanned the entire number line in size order. In *Sort: Nonlinear* boxes were arranged in a 2-by-2 array with the extra box in the middle. (See Figure 2.) Participants completed 7 practice Sort trials to further orient them to the range of boxes in which they would later categorize numbers: the two endpoints of the number line (i.e. 0 and 100) and the 5 box midpoints (i.e. 10 [half way between 0 and 19], 30 [half way between 20 and 39], and so on, to get 50, 70, 90).

The third and most substantial part of the activity was a series of 16 trials during which participants did a sort trial immediately followed by an estimation trial. *Estimate*, immediately followed each *Sort* task. (See Figure 2.) *Estimate* contained the same features and components found in the pretest activity, MANL, except for the addition of an adjustable range, or the *user-defined range* (UDR) (Creighan, 2014). The UDR replaced the 5-unit horizontal range participants used to indicate their estimate in session 1 and was adjustable in length. It allowed participants to use the range length to indicate their level of confidence in their estimates. If very confident, they could use a very narrow range. If not confident, they could enlarge the range. The UDR was controlled by moving the mouse or the 'Left' and 'Right' arrow keys and was made wider or narrower by manipulating its endpoints. Because the UDR did not have maximum or minimum restrictions, users could make the range as large as the number line or as small as a point. The default size of the UDR was 5 units. The narrator gave the estimation instructions as follows: "The number line is from 0 to n. (*Target value appears*). Where does n belong on the number line? Show how sure you are."

Participants estimated the same numbers as in the sort portion of the activity. The numbers were presented in the same randomized order as *Sort*. The numbers selected for the trials provide a balanced set of data points across the number line and are as follows: 3, 9, 17, 24, 31, 36, 45, 49, 52, 58, 64, 72, 78, 85, 93, 95. Three numbers were selected for each range of 20 numbers, or quintile, except for the middle quintile, which had four numbers – two on either side of the midpoint, 50. Participants saw one number at a time. While the 16 trials were fewer than was used in MANL and in other number line studies (e.g. Opfer & Siegler, 2007), because MASE was a two-part activity it was as though the number of trials was doubled. Further, an informal pilot study suggested fatigue and boredom were an issue with higher numbers of trials, particularly for the younger participants in this cross-sectional study.

Chapter 3

Results

Analysis

When considering number line estimates, one can think about the bias of estimates as a magnitude of error, namely percent absolute error (PAE) (Booth & Siegler, 2006; Opfer & Thompson, 2008), or about the precision of estimates as the standard deviation. To determine the percent absolute error, the absolute value of the difference of the estimate and the actual magnitude is divided by the range:

$$PAE=\frac{|Estimate-Target\ Value|}{Range}\times 100$$ (Siegler & Ramani, 2009).

The PAE gives a nuanced view of number line estimation behavior by providing a specific *magnitude* of error, not only a binary, like correct versus incorrect. It is characteristic of the PAE that as number line estimation skill improves PAE decreases.

The standard deviation of the estimates is a measure of how spread the estimates are on the number line. A large standard deviation means that the estimates are widely dispersed. A small standard deviation means that the estimates are close together.

To analyze the PAE and the standard deviation I conducted a systematic examination of the means. I used t-tests when comparing a pair of means. I used analyses of variance (ANOVAs) to determine if a group of means significantly vary. Along with the ANOVA, I used Tukey's honest significant difference (HSD) tests in order to find the pairs of means that significantly differ while adjusting for doing multiple comparisons. Finally, I used hierarchical linear models (HLMs) to examine the difference in PAE on an individual level. Typical regression models assume independent errors and that "regression coefficients apply equally to all contexts" (Luke, 2004, 7). Because sessions are nested within participants, error cannot be assumed to be independent and coefficients may differ across sessions.

Initial analysis of the PAE showed that the data was non-normal (see Figures 3-4). As the PAE is not normally distributed I used a natural log transformation to normalize the data prior to carrying out analyses (see Figures 5-6). The logit mean PAE is used for all analysis.

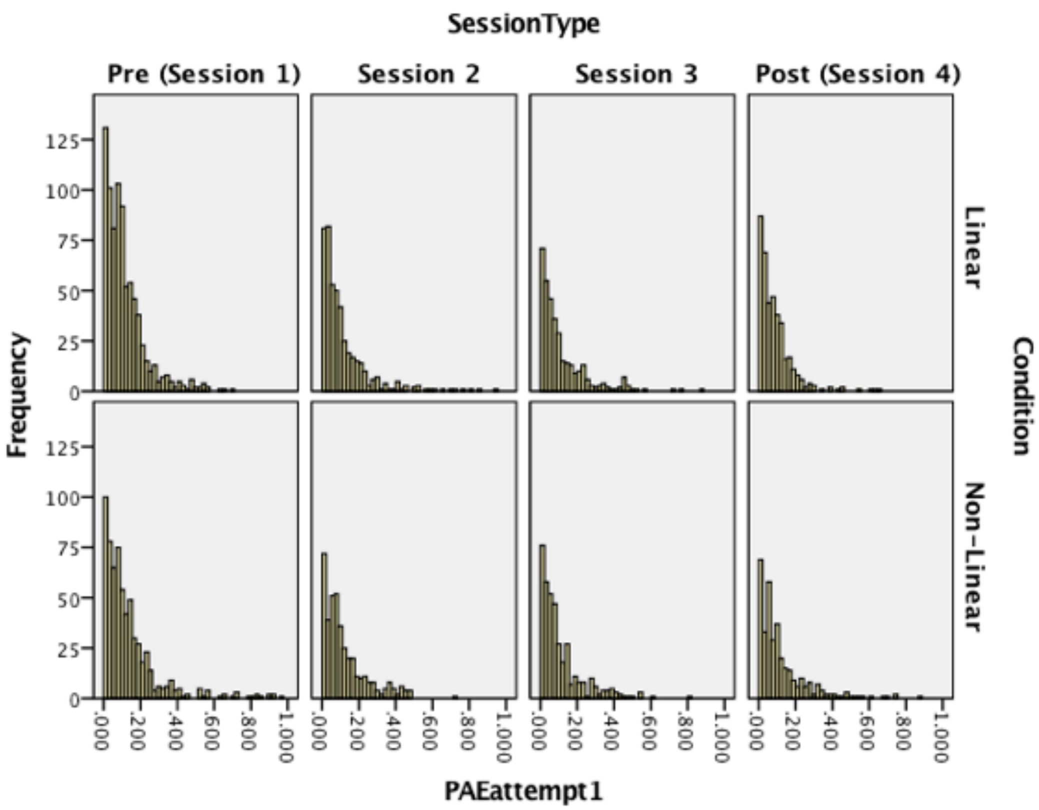


Figure 3. Non-transformed percent absolute error (PAE) by phase by condition.

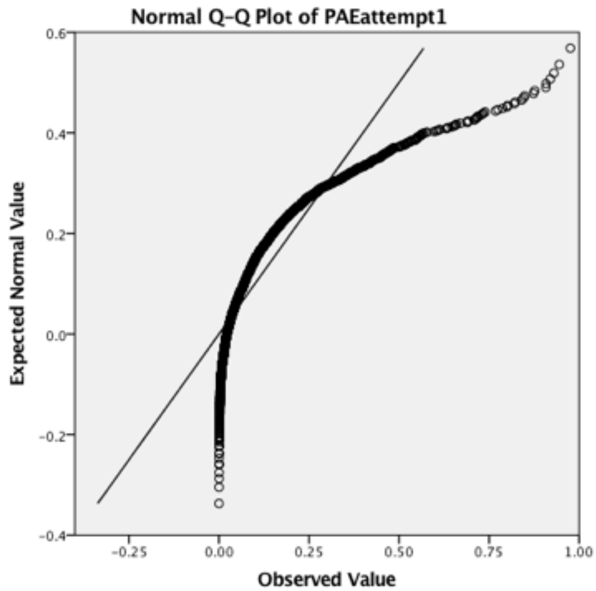


Figure 4. Quartile-Quartile plot of untransformed PAE data

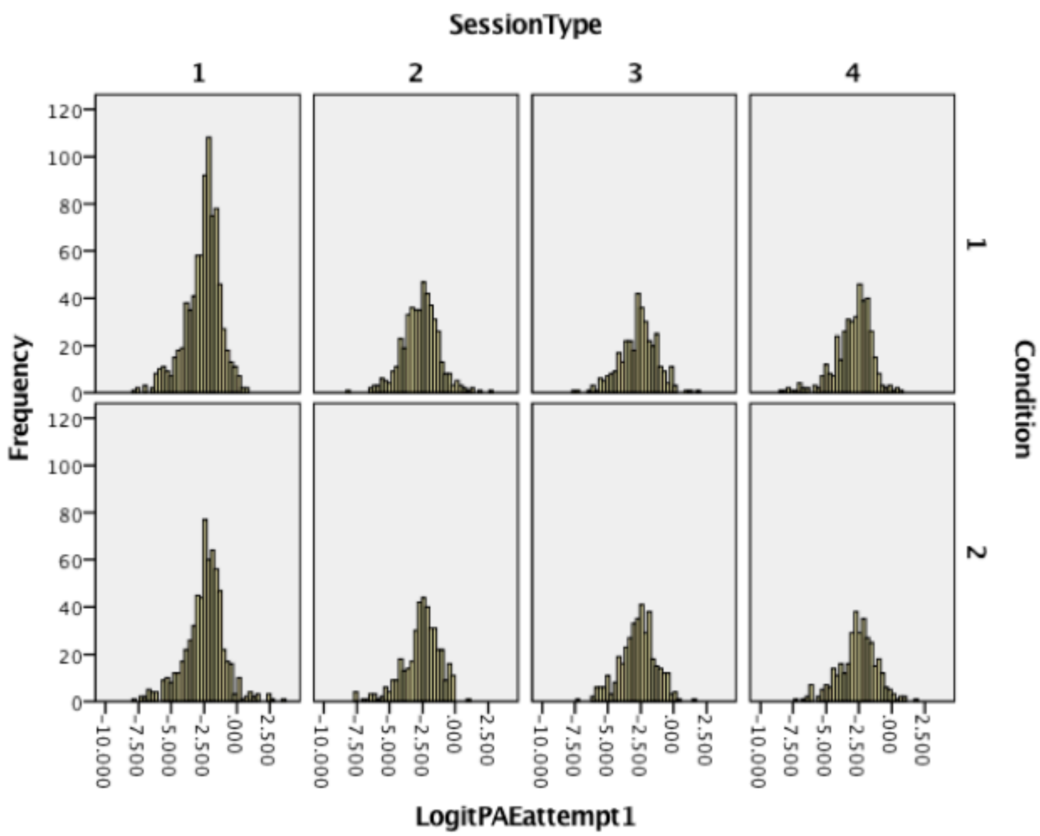


Figure 5. Log transformed percent absolute error by phase by condition

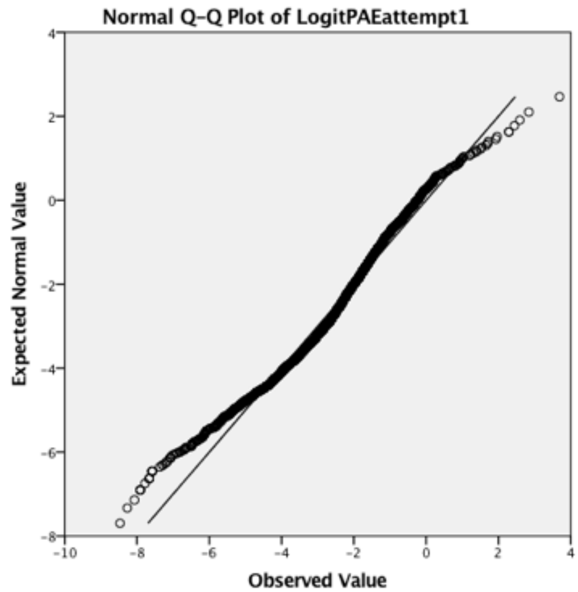


Figure 6. Quartile-Quartile plot of transformed PAE data

In service of answering the broader question, *Can a computer-based number categorization task and a computer-based number line estimation task improve children's number line performance?,* the following questions are addressed:

1) *How do children perform on a computer-based number* ***categorization*** *task?* 2) *How do children perform on a computer-based* ***number line estimation*** *task?* 3) *How does number line estimation performance change over time?* 4) *How does number categorization affect number line estimation?* 5) *How do individual differences affect number categorization and number line estimation?*

The first two questions provide a descriptive look at the data. The third question is a comparative question that examines the data with respect to time. The fourth question compares the linear and nonlinear groups. And the fifth and final question takes into account individual differences while examining performance over time and across conditions.

1) How do children perform on a computer-based number ***categorization*** *task?*

Examination of the bar graphs (Figure 7) of the mean percentage of correct categorizations from session 1, suggests that differences occur between grades. The mean percents increase as grade-level increases (Table 3). A one-way analysis of variance (ANOVA) that has mean percentage of correct categorizations as the outcome variable and grade-level as the predictor variable shows that the grade-level means do significantly differ ($F(2, 45) = 8.185$, $p = .001$, $\eta^2 = 0.27$). In order to know specifically which means differed I conducted a post-hoc analysis using Tukey's honest significant difference (HSD) test. (See Table 4.) For mean percent correct, Grade 1 differs from Grades 2 and 3. The first grade participants, on average, sorted fewer numbers correctly than the participants in Grades 2 and 3. As hypothesized older grades outperform younger grades in number categorization.

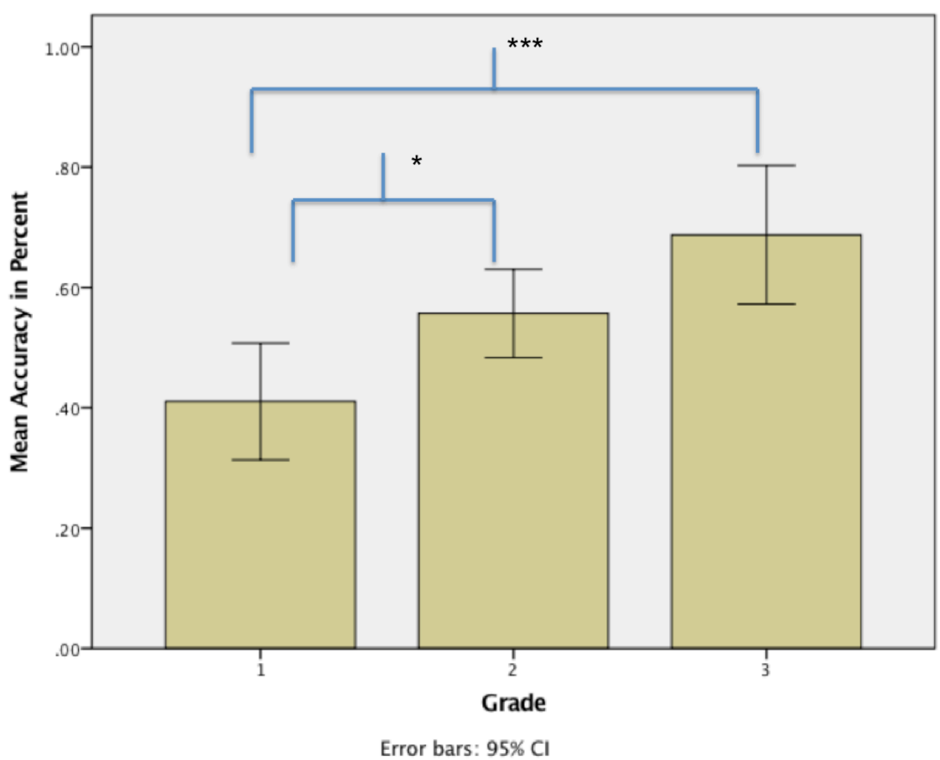


Figure 7. Mean accuracy for number categorization in percent for session 1. *p<.05; ***p<.001.

Table 3 Mean percentages for number categorization accuracy

Grade	Means
1	41.05%

2	55.68%
3	68.75%

Table 4 Post-hoc comparisons of grade-level number categorization percent accuracy

Grades	Mean difference	p-value
1 vs 2	-0.14635	.030
1 vs 3	-0.27703	.001
2 vs 3	-0.13068	.143

In addition to understanding the differences in performance on the number categorization task, it is important to examine the differences in performance on the number line estimation task to better understand from where participants are starting. I compared the mean PAE for the pretest across grade-levels to determine if participants in different grades perform differently on the computer-based number line estimation task and report the results in the following question.

2) How do children perform on a computer-based ***number line estimation*** *task?*

From the boxplots (Figure 8) we see lower medians and less variation as grade increases. The variability in Grade 1 is aligned with Barth, Starr, & Sullivan (2009) assessment that in this age range knowledge of numbers up to 100 is highly variable.

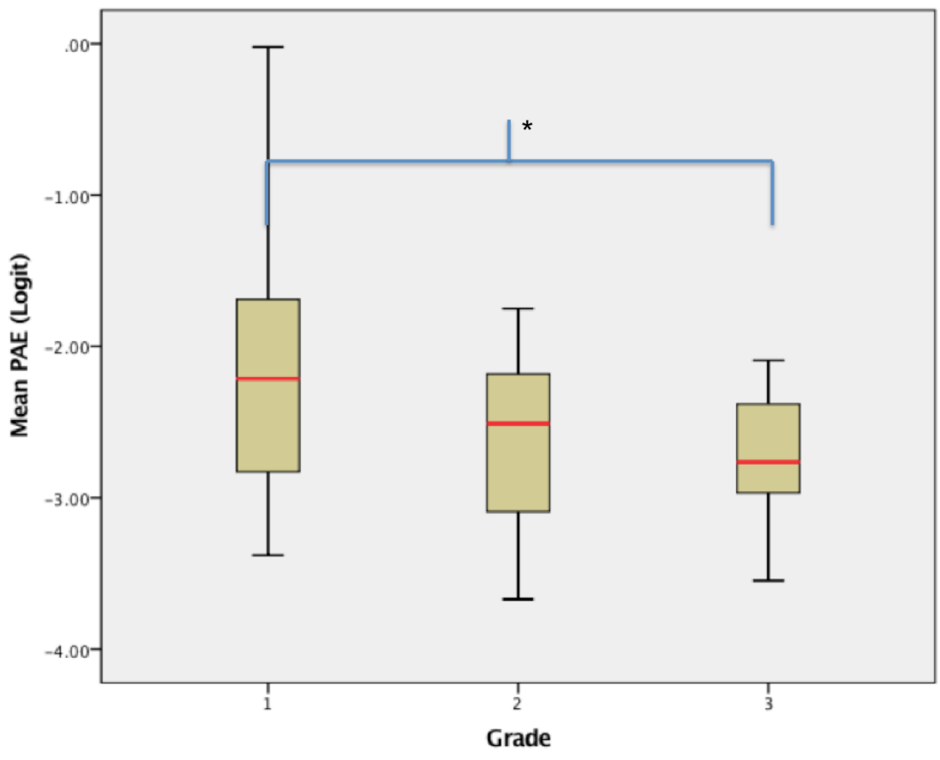


Figure 8. Mean PAE (Logit) for number line estimation. The red line represents the median PAE. *p<.05.

A one-way ANOVA with mean PAE as the outcome variable and grade as the predictor variable suggests that mean PAEs do significantly differ ($F(2, 55) = 4.005$, $p = .024$, $\eta^2 = 0.13$). The mean PAEs for each grade are shown in table 5: mean increases as grade increases with Grade 3 having the lowest mean PAE.

Table 5 Mean PAE (Logit) for number line estimation

Grade	Means
1	-2.1315968
2	-2.6157668
3	-2.7586448

A post-hoc test compared each grade's mean PAE (Table 6). Again, as with the categorization task, grade-level differences are present. Grade 3 has significantly lower mean PAE than Grade 1. Grade 2's mean PAE is trending toward being significantly lower than Grade 1's mean PAE.

Table 6 Post-hoc test comparing mean PAEs

Grades	Mean	p-value

	difference	
1 vs 2	0.484169945	.052
1 vs 3	0.62704796	.047
2 vs 3	0.142878018	.831

3) How does number line estimation performance change over time?

To see how number line estimation performance changes over time the session data for the non-treatment (i.e. nonlinear) group is helpful. From the session data of the nonlinear group a U-shaped trend is noticeable (Figure 9); mean PAE first decreases than increases to its original starting point. From the graph it appears there is not an overall change in error over time. To test if the sessions significantly vary from one another would require an analysis method that can accommodate the sessions being nested within participants. Overall, participants are performing about the same on all sessions without the linear treatment.

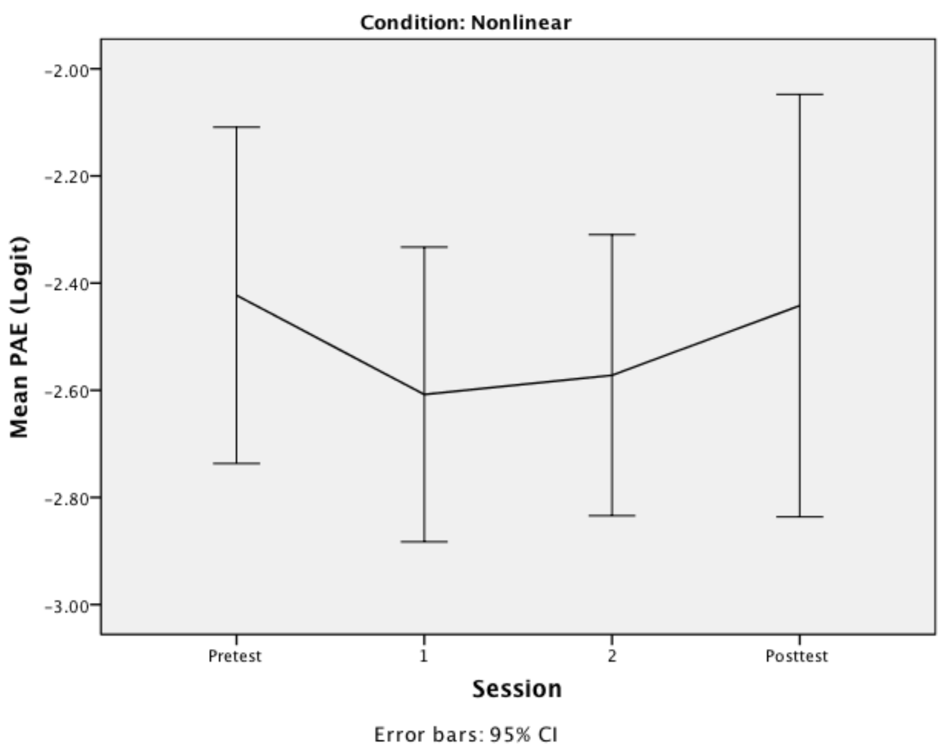


Figure 9. Mean PAE (Logit) for each phase for the non-treatment (i.e. nonlinear) group.

4) How does number categorization affect number line estimation?

Error is comparable between the linear and nonlinear groups for the first three phases. (See Figure 10.) The two groups have comparable mean PAEs for all the phases of the study until the posttest where the groups diverge. The linear group has less mean PAE (t(27) = -2.4421, p = .029, d = 0.86), or is performing better, than the nonlinear group. (See Table 7.)

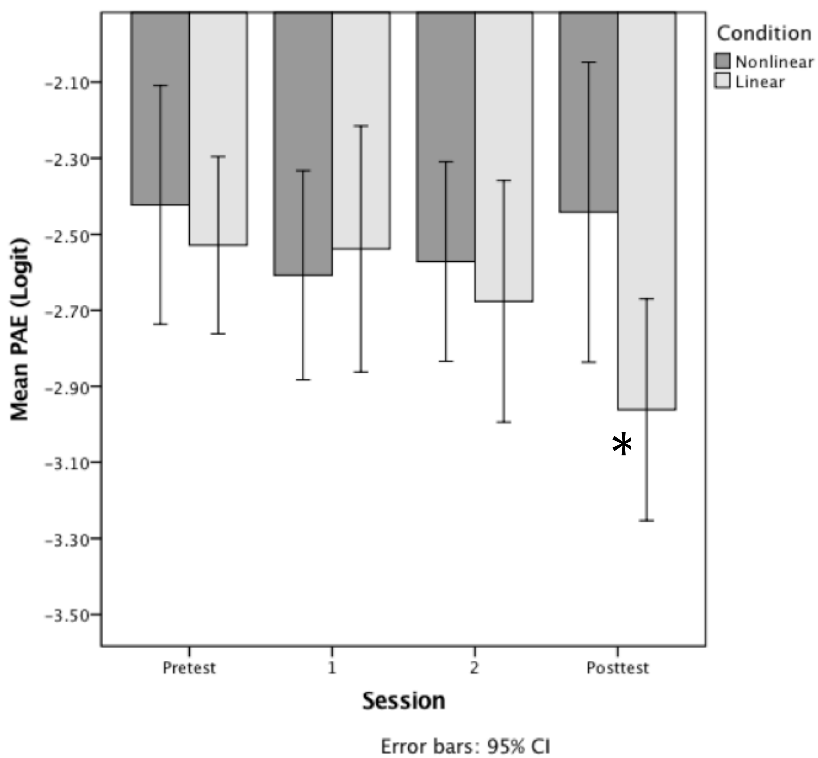


Figure 10. Mean PAE (Logit) over phases of study.
*p<.05.

Table 7 Comparison of mean PAE by session across conditions.

Session	Nonlinear	Linear	t	df	p-value
1	-2.4228393	-2.5289790	-0.556	56	.580
2	-2.6078034	-2.5386248	0.334	46	.740
3	-2.5719528	-2.6766666	-0.534	39	.597
4	-2.4421	-2.9614	-2.304	27	.029

In comparing the standard deviation of the linear group to the nonlinear group over the four sessions (see Figure 11), the two groups do not differ much (Table 8) except in session 1, where the linear group has a substantially smaller standard deviation (t(46) = 2.181, p =.034, d =

0.63). Or stated differently, the linear group has less dispersed estimates in session 1, the first session with the scaffold.

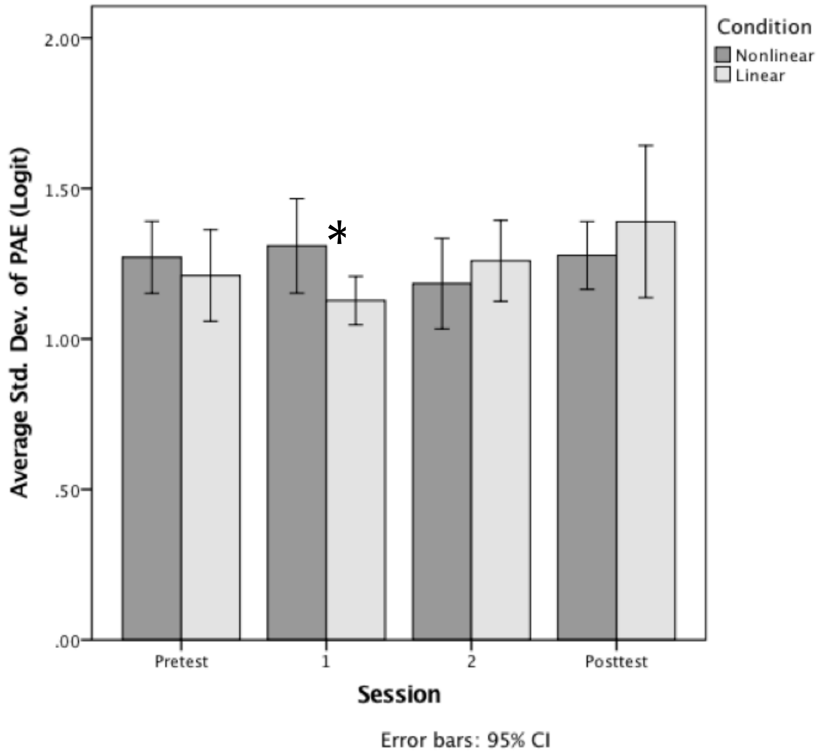


Figure 11. Standard deviation of mean PAE (Logit) over phases of study.
*p<.05.

Table 8 t-tests comparing the standard deviation of the mean PAE

Session	Nonlinear	Linear	t	df	p-value
1	1.27092306	1.21061416	0.638	56	.526
2	1.30884454	1.12736868	2.181	46	.034
3	1.18359386	1.25941145	-0.783	39	.439
4	1.27730901	1.38947161	-0.850	27	.403

When examined simultaneously, the error and standard deviation of the estimates do not increase and decrease in a similar pattern. In the treatment condition, the linear group, first a decrease in standard deviation occurs, followed by an increase in the standard deviation but an a decrease in PAE. And the results of the comparisons hold regardless of whether looking at the full sample or only at the participants who complete all four phases (Table 9). Thus, the results are likely not an effect of attrition.

Table 9. Mean PAE and standard deviation by condition across study phases for the entire sample and for the group that completed all four phases.

	Pretest	Session 1	Session 2	Posttest
Full Sample				
Nonlinear				
Mean PAE	-2.4228	-2.6078	-2.5720	-2.4421
Mean SD	1.2709	1.3088	1.1836	1.2773
Linear				
Mean PAE	-2.5290	-2.5386	-2.6767	-2.9614*
Mean SD	1.2106	1.1274*	1.2594	1.3895
Completed				
Nonlinear				
Mean PAE	-2.4033	-2.7096	-2.6423	-2.4421
Mean SD	1.3599	1.3943	1.0999	1.2773
(n=14)				
Linear				
Mean PAE	-2.6388	-2.7465	-2.9147	-2.9614*
Mean SD	1.2447	1.1020***	1.2035	1.3895
(n=15)				

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

The differences between the linear and nonlinear group are also apparent when looking at performance by number line quintiles. Number line estimation performance varies by number (Creighan, 2014) with smaller numbers being easier to estimate than larger numbers (Opfer & Siegler, 2007). Examining number line quintiles allows investigation of bands of common numbers and is a more parsimonious approach than exploring differences by each individual number. Figure 12 shows number line estimation error by each quintile for the four phases for the linear and nonlinear groups. The linear and nonlinear groups perform comparably well on the five quintiles at each phase until the fourth phase, the phase in which the scaffold is removed. In the posttest the linear group estimates better than the nonlinear group in quintile 1 (0-19) and quintile 2 (20-39).

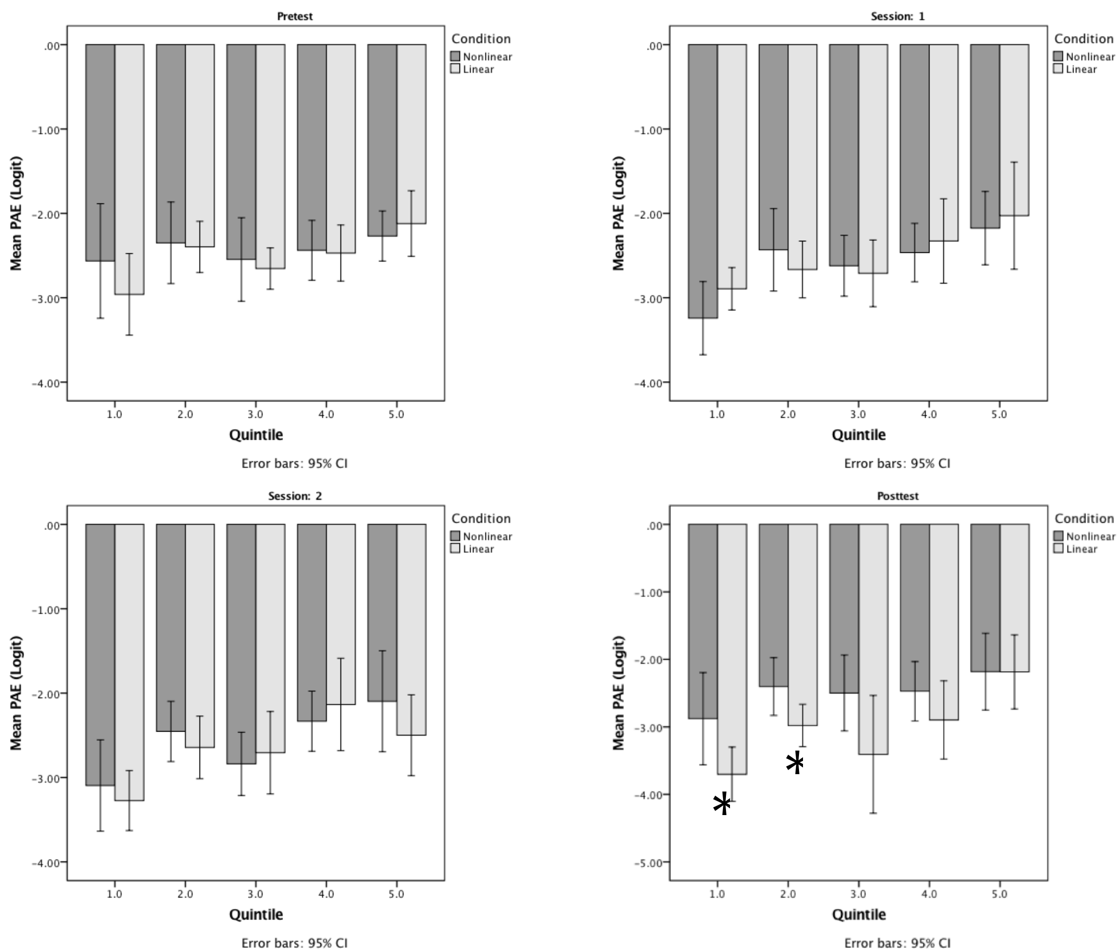


Figure 12. Mean PAE by quintiles by study phase.
*p<.05.

5) How do individual differences affect number categorization and number line estimation?

As predicted initial individual differences in numeracy as measured by a magnitude comparison task dictated student performance throughout the study. The participants with lower numeracy, those who scored lower on the digit comparison task at the start who were in the linear group, had significantly less error on their estimates than the nonlinear low numeracy participants, whereas the linear and nonlinear groups with high numeracy at the study's outset did not differ (see Figure 13).

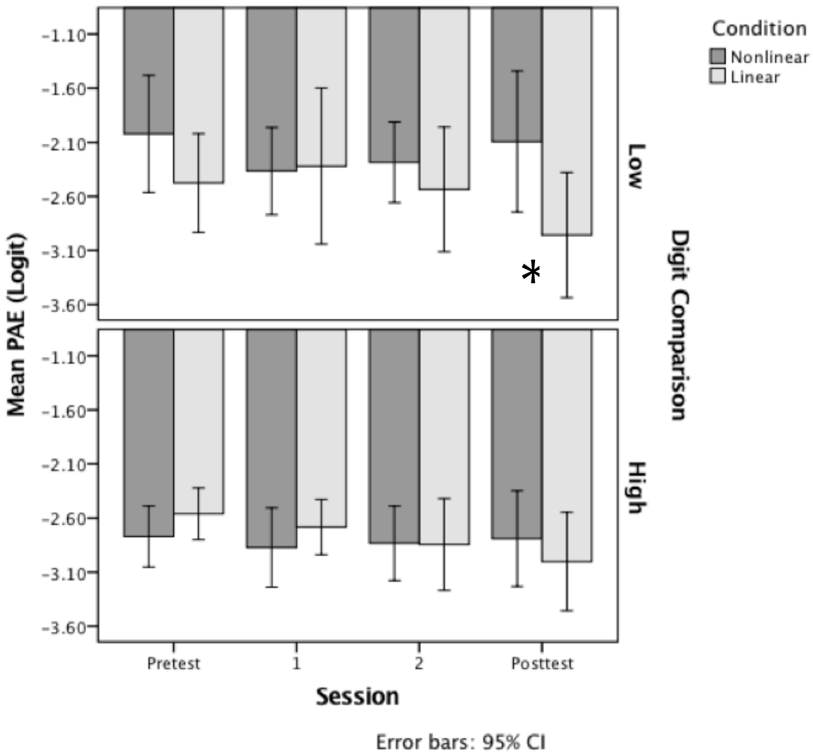


Figure 13. Comparing number line estimation mean PAE by condition in low- and high-scoring digit comparison groups. Of the participants who scored low on the digit comparison task, the linear group did significantly better than the nonlinear group at posttest ($p = .032$, $d = 1.298$). *$p<.05$.

Growth Curves

Growth and learning occur within and across the phases of the study. Further, to understand growth in the scope of number line estimation skills it is necessary to take into account all the variables that could account for a change in PAE. To identify patterns of growth in mean PAE over time I used a 2-level growth curve model that takes into account the individual differences that affect number line estimation. Mean PAE was the outcome variable (i.e. MEANLOGITPAE). The time variable (i.e. SESSPOST) was centered at phase 4 of the study, so phase 4, the point at which there was the most data and at which the greatest differences were expected, was the intercept of the model. Time (i.e. SESSPOST) was added at level 1 of the model.

I incorporated various predictor variables to control for their role in predicting mean PAE, the most relevant of which was student's score on the pretest MANL task. All level 2 variables were grand-mean centered so that the intercept could be interpreted as the average outcome for the average student in the control condition. The mixed model is shown below:

$$MEANLOGITPAE_{ij} = \gamma_{00} + \gamma_{01}*PRETEST_j + \gamma_{02}*TREATMENT_j + \gamma_{03}*SCHOOL_j + \gamma_{04}*AGE_j + \gamma_{05}*SEX_j + \gamma_{06}*GRADE2_j + \gamma_{07}*GRADE3_j$$

$$+ \gamma_{10}*SESSPOST_{ij} + \gamma_{11}*PRETEST_j*SESSPOST_{ij} + \gamma_{12}*TREATMENT_j*SESSPOST_{ij} + \gamma_{13}*SCHOOL_j*SESSPOST_{ij} + \gamma_{14}*AGE_j*SESSPOST_{ij} + \gamma_{15}*SEX_j*SESSPOST_{ij} + \gamma_{16}*GRADE2_j*SESSPOST_{ij} + \gamma_{17}*GRADE3_j*SESSPOST_{ij}$$

$$+ u_{0j} + r_{ij}$$

Table 10 provides all the variables in the model along with their definitions.

Table 10. Terms in the main hierarchical linear model

Variable	Definition
MEANLOGITPAE	Outcome variable; mean percent absolute error in logits
PRETEST	Pretest score: Mean PAE on the MANL task
TREATMENT	Whether the participant was in the treatment group or not; linear versus nonlinear
SCHOOL	Which school the participant attended
AGE	Participant's age
SEX	Participant's sex
GRADE2	Whether the participant was in Grade 2 or not
GRADE3	Whether the participant was in Grade 3 or not
SESSPOST	Session centered at phase 4 of the study

Table 11 shows the results of the final main model and Figure 14 shows the average relationship over time between the treatment (i.e. linear) and control (i.e. nonlinear) groups.

Table 11 The HLM results for the main effects model. PRETEST has been grand mean centered

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value

For INTRCPT1, β_0					
INTRCPT2, γ_{00}	-2.393747	0.134650	-17.778	41	<.001
PRETEST, γ_{01}	0.270336	0.119886	-2.255	41	.030
TREATMENT, γ_{02}	-0.293460	0.164067	-1.789	41	.081
For SESSPOST slope, β_1					
INTRCPT2, γ_{10}	0.157280	0.071593	2.197	61	.032
PRETEST, γ_{11}	-0.183659	0.060780	-3.022	61	.004
TREATMENT, γ_{12}	-0.246264	0.083590	-2.946	61	.005

Note: Model also adjusted for School, Age, Sex, and Grade all of which were grand mean centered.

Table 11 indicates that while the treatment group had logit (PAE) -.29 points lower than the control group at Phase 4, posttest, this difference was not statistically significant (p = .081). However, the change in logit (PAE) over time does differ in the treatment and control groups (p = .005). In particular, while the logit (PAE) was found to *increase* over time (0.15 for each time point), in the treatment group the logit (PAE) *decreased* over time (0.15 – 0.24 = -0.09 for each time point). In other words the linear group is growing in number line estimation skills at a faster rate than the nonlinear group.

In Figure 14, the graphic representation of the model reflects the results shown in Table 11. Additionally, the graph suggests that initially the nonlinear group has a lower mean PAE. Over time the groups switch positions with the linear group having the lower mean PAE and that relationship stays constant. The figure also suggests that – if the trends in both groups continued – the differences between the treatment and control groups would grow over time. While this is an extrapolation, it generates hypotheses regarding directions for future research.

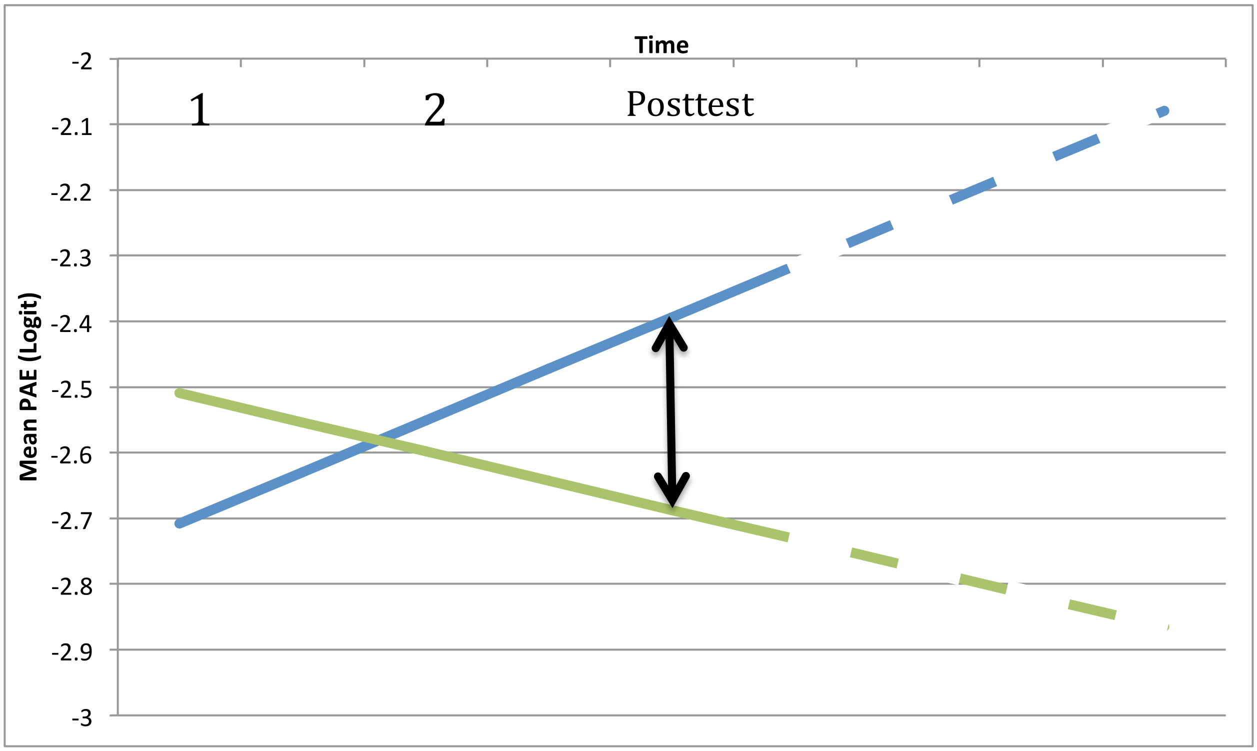


Figure 14. The main effects model in graphical form. The blue line is the nonlinear group. The green line is the linear group. The black double-sided arrow indicates the difference in mean PAE at posttest. The dashed lines represent an extrapolation of the trend if more time points were observed. The slopes of the two lines are significantly different ($p = .005$).

Treatment Effects by Pretest Ability

As stated in Chapter 1, pretest differences have been important in number line estimation studies (Creighan, 2014; Ramani & Siegler 2008; Siegler & Ramani, 2009). Ramani and Siegler (2008) found that those participants who had low pretest scores performed at the low end. Another study (Siegler & Ramani, 2009) found that participants with low pretest scores made the largest gains in number line estimation skills. In order to assess if there are pretest differences in this study participants must be separated by their pretest scores and analyzed.

In order to study the effect of the pretest, I used a median-split to group participants by their performance on the MANL pretest task and added this categorical variable (i.e. PREHIGH) to the model to explore if there were differences in growth curves for the participants who scored low and high on the MANL pretest task. The models for the low and high pretest groups are identical and also identical to the main model shown above. For convenience the mixed model is presented again below:

$MEANLOGITPAE_{ij} = \gamma_{00} + \gamma_{01}*PRETEST_j + \gamma_{02}*TREATMENT_j + \gamma_{03}*SCHOOL_j +$

$\gamma_{04}*AGE_j + \gamma_{05}*SEX_j + \gamma_{06}*GRADE2_j + \gamma_{07}*GRADE3_j$

$+ \gamma_{10}*SESSPOST_{ij} + \gamma_{11}*PRETEST_j*SESSPOST_{ij} + \gamma_{12}*TREATMENT_j*SESSPOST_{ij}+$

$\gamma_{13}*SCHOOL_j*SESSPOST_{ij}+ \gamma_{14}*AGE_j*SESSPOST_{ij}+ \gamma_{15}*SEX_j*SESSPOST_{ij}+$

$\gamma_{16}*GRADE2_j*SESSPOST_{ij}+ \gamma_{17}*GRADE3_j*SESSPOST_{ij}$

$+ u_{0j}+ r_{ij}$

The variables are defined in Table 10. Tables 12 and 13 give the model coefficients for the low and the high pretest groups. The treatment (i.e. linear vs. nonlinear) is not significant at the intercept for either the low or high group. Thus, at Phase 4, the posttest, the linear and nonlinear groups do not differ in mean PAE regardless of if they had a low or high MANL pretest score. However the treatment variable is significant in the slope of the low pretest score group ($p = .007$), but not the high pretest score group. Focusing only on the low pretest score group: the negative coefficient on TREATMENT indicates that the linear group's mean PAEs are getting more negative over time (0.23 - 0.55 = -0.32) and at a rate that is significantly different ($p = .007$) from the nonlinear group. Growth in the linear group is faster than growth in the nonlinear group.

Table 12. The HLM results for the low pretest score group.

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	-2.034320	0.264534	-7.690	18	<.001
PRETEST, γ_{01}	0.411394	0.279231	1.473	18	.158
TREATMENT, γ_{02}	-0.602098	0.367254	-1.639	18	.118
For SESSPOST slope, β_1					

INTRCPT2, γ_{10}	0.234110	0.137915	1.697	21	.104
PRETEST, γ_{11}	-0.155064	0.134243	-1.155	21	.261
TREATMENT, γ_{12}	-0.551187	0.184448	-2.988	21	.007

Note: Model adjusted for School, Age, Sex, and Grade all of which were grand mean centered.

Table 13 The HLM results for the high pretest score group.

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	-2.632218	0.170269	-15.459	15	<.001
PRETEST, γ_{01}	0.511361	0.373334	1.370	15	.191
TREATMENT, γ_{02}	-0.377875	0.206867	-1.827	15	.088
For SESSPOST slope, β_1					
INTRCPT2, γ_{10}	0.135419	0.082647	1.570	32	.126
PRETEST, γ_{11}	0.057464	0.182074	0.316	32	.754
TREATMENT, γ_{12}	-0.161903	0.099161	-1.633	32	.112

Note: Model adjusted for School, Age, Sex, and Grade all of which were grand mean centered.

Figure 15 shows the graphs for the low and high pretest groups. In the graphs the low and high pretest groups follow the same growth pattern that was in the main model: the nonlinear groups start with a lower mean PAE than the linear groups and gradually the linear groups' mean PAE becomes lower. Over time the linear groups perform better, although not significantly, similar to the performance pattern seen in the main linear model. The low group's growth is a steeper downward slope than that of the high group, which is almost flat. The low group's growth mirrors that which was seen in the main effect model. Most of the linear group's significant growth rate from the main model is being accounted for by the growth rate of the low linear group, since the high linear group is hardly contributing. This result provides hypotheses for future research into if participants with low pretest scores decrease their number line estimation error faster than participants with high pretest scores.

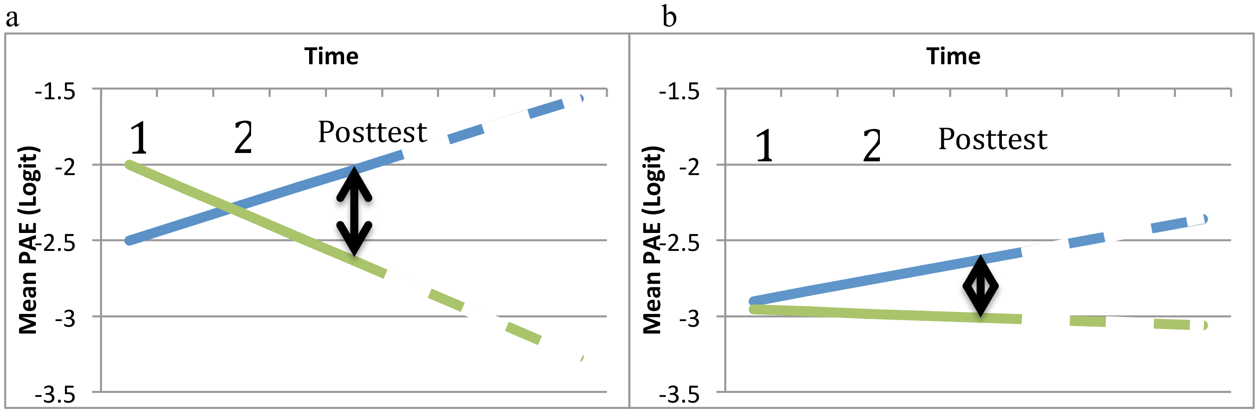


Figure 15. The model in graphical form for participants who scored low (15a) and high (15b) on the pretest MANL task. The blue lines are the nonlinear groups. The green lines are the linear groups. The black double-sided arrow indicates the difference in mean PAE at Phase 4, the posttest. The dashed lines represent an extrapolation of the trend if more time points were observed.

Finally, it was important to test if an interaction existed between the low and high pretest score groups. The mixed model for testing that interaction is below:

$$MEANLOGITPAE_{ij} = \gamma_{00} + \gamma_{01}*PREHIGH_j + \gamma_{02}*TREATMENT_j + \gamma_{03}*SCHOOL_j +$$

$$\gamma_{04}*AGE_j + \gamma_{05}*SEX_j + \gamma_{06}*GRADE2_j + \gamma_{07}*GRADE3_j + \gamma_{08}* TRTMNTxPREHIGH_j +$$

$$\gamma_{10}*SESSPOST_{ij} + \gamma_{11}*PRETEST_j*SESSPOST_{ij} + \gamma_{12}*TREATMENT_j*SESSPOST_{ij}+$$

$$\gamma_{13}*SCHOOL_j*SESSPOST_{ij}+ \gamma_{14}*AGE_j*SESSPOST_{ij}+ \gamma_{15}*SEX_j*SESSPOST_{ij}+$$

$$\gamma_{16}*GRADE2_j*SESSPOST_{ij}+ \gamma_{17}*GRADE3_j*SESSPOST_{ij}\ \gamma_{13}*$$

$$TRTMNTxPREHIGH_j*SESSPOST_{ij}$$

$$+ u_{0j}+ r_{ij} .$$

The terms in the model are defined in Table 14.

Table 14 Terms in the hierarchical linear model in which pretest MANL task performance divides participants into a high and a low group.

Variable	Definition
MEANLOGITPAE	Outcome variable; mean percent absolute error in logits
PREHIGH	Whether the participant was in the high pretest score group based on a median split

TREATMENT	Whether the participant was in the treatment group or not; linear versus nonlinear
SCHOOL	Which school the participant attended
AGE	Participant's age
SEX	Participant's sex
GRADE2	Whether the participant was in Grade 2 or not
GRADE3	Whether the participant was in Grade 3 or not
SESSPOST	Session centered at phase 4 of the study
TRTMNTxPREHIGH	Interaction term for treatment and the digit comparison task groups

Table 15 HLM for participants who scored low and high on the pretest MANL task.

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	-2.235999	0.195685	-11.427	40	<.001
PREHIGH, γ_{01}	-0.231241	0.273146	-0.847	40	.402
TREATMENT, γ_{02}	-0.537797	0.293542	-1.832	40	.074
TRTMNTxPREHIGH, γ_{08}	0.230537	0.396148	0.582	40	.564
For SESSPOST slope, β_1					
INTRCPT2, γ_{10}	0.086194	0.096629	0.892	60	.376
PREHIGH, γ_{11}	0.121861	0.126402	0.964	60	.339
TREATMENT, γ_{12}	-0.400071	0.141021	-2.837	60	.006
TRTMNTxPREHIGH, γ_{18}	0.277216	0.184614	1.502	60	.138

Note: Model also adjusted for School, Age, Sex, and Grade all of which were grand mean centered.

Table 15 shows the coefficients for the interaction model with participants broken into groups based on performance on the pretest MANL task. The coefficients for PREHIGH and for the interaction variable (i.e. TRTMNTxPREHIGH) are not significant at the intercept or the slope of the model (all $p > .1$). These results suggest that the high and low groups do not have different mean PAEs. Also, since the interaction variable is not significant, the growth pattern in the low group is the same as the growth pattern in the high group.

Chapter 4

Discussion

Number line estimation performance reflects children's understanding of magnitude and their understanding of the relationships between numbers. Number line estimation is associated with performance on other mathematical tasks, such as functional numeracy measures (Geary et. al., 2013), arithmetic (Booth & Siegler, 2008; Cowan & Powell, 2014) and standardized mathematics assessments (Ashcraft & Moore, 2012; Booth & Siegler, 2006; Siegler & Booth, 2004; Sasanguie, De Smedt, Defever, Reynvoet, 2012). Since number line estimation is associated with performance on other mathematical tasks, improving number line estimation skills may improve performance on the tasks with which number line estimation is associated. Regardless of the possibility that number line estimation can improve performance on other mathematical tasks, it is a strong measure and tool in itself. It is important to improve number line estimation skills.

The present study sought to improve number line estimation skills through a computerized magnitude-based intervention. Participants categorized numbers according to their subjective view of the magnitude of those numbers. They categorized numbers according to magnitude linearity or in a nonlinear configuration. I explored five questions in the study to better understand the relationship between computer-based categorization and number line estimation. According to research, one would expect there to be age or grade-level differences evident in categorization accuracy (Laski & Siegler, 2007) and in number line estimation error (Opfer & Siegler, 2007) with the older participants outperforming the younger participants. Further, I expected participants would improve in number line estimation error over time, from session to session, due to Creighan's (2014) work. Additionally, I hypothesized that the participants in the linear group would improve their number line estimation error more than the participants in the nonlinear group because of the *representational mapping hypothesis* (Siegler & Ramani, 2009). Finally, I hypothesized that participants' individual performance on numeracy and number identification measures would be related to their number line estimation error and the gains they made. Let us examine each question and the associated results.

Question 1: How do children of different ages perform on the number categorization task?

Older participants performed best on the number categorization task, similar to Laski and Siegler's (2009) results. Participants in Grade 3 categorized more accurately than Grade 1 (p=0.001) and Grade 2 also categorized better than Grade 1 (p=0.030). Laski and Siegler's (2007) participants were in grades Kindergarten – 2^{nd} and their categorization task involved placing numbers into physical baskets. The present study had an older participant age range and used a computerized categorization task. However, the similarities in grade-level differences lend validity to the computer-based version of the number categorization task.

Question 2: How do children perform on a computer-based number line estimation task?

There was a positive relationship between grade and performance on the number line estimation task with older participants performing better than younger participants. Grade 3 participants estimated with significantly less error than Grade 1 participants (p=0.047). And while Grade 2 participants did not estimate significantly better than Grade 1 participants (p=0.052), their performance was trending in that direction. These results mirror previous cross-sectional studies (e.g. Siegler & Booth, 2004) and lend validity to this computer-based number line estimation task

Question 3: How does number line performance change over time?

The non-treatment (i.e. nonlinear) group's performance was in a u-shaped pattern:

number line estimation error reduced than increased back to its original starting point for no net change in error, although the changes in error were not significant. From a previous study (Creighan, 2014) I expected that number line estimation skills would improve over time simply from exposure to the number line estimation task. A decrease in percent absolute error (PAE) would demonstrate an improvement in number line estimation skills. This was not the case. The nonlinear number categorization task in this study was not present in Creighan's (2014) study and the task may have interfered with improvement occurring in number line estimation error. While number line estimation error did not improve, it is important to note that the nonlinear categorization participants did not worsen either.

Question 4: How is computer-based number categorization associated with computer-based number line estimation?

Exposure to the linear boxes scaffold was associated with an improvement in number line estimation error, whereas exposure to the nonlinear boxes scaffold was not associated with an improvement in number line estimation error. At pretest the linear and nonlinear groups did not differ in number line estimation error. However, by the posttest, on average, the linear participants had significantly lower PAE than the nonlinear participants ($t(27)=-2.9614$, $p<0.05$). These findings suggest that linearity is an important aspect to increasing number line estimation performance. The phenomenon with the linear boxes was similar to what Laski and Siegler (2007) found, namely that participants who categorized numbers by magnitude improved their number line estimation linearity without training on number line estimation. One major difference between Laski and Siegler's (2007) study and this study is that in the present study participants did not receive feedback when sorting. Laski and Siegler (2007) implicate feedback as being the mechanism for change in number line estimation linearity. However, since no feedback was given during the present categorization intervention it was unlikely that feedback was the sole cause for change in number line estimation performance in Laski and Siegler (2007). Alternate theories for the improvement in number line estimation are offered later in this chapter.

The difference in number line estimation error between the linear and nonlinear groups at posttest was not evident for all segments of the number line. Error differed by quintiles, with the linear group performing better in quintiles 1 ($t(27)=2.277$, $p=0.031$) and 2 ($t(27)=2.361$, $p=0.026$) and not performing differently for quintiles 3, 4, and 5 (all $p>0.075$). Potential reasons for the difference in quintile performance will be discussed later in the chapter.

Question 5: How is individual difference related to number line estimation performance?

Of the pretreatment measures (i.e. number identification, dot comparison, and digit comparison), only the digit comparison task differentiated participants. I used a median split to stratify participants into a low and high group. Participants with low scores on the digit comparison task in the linear condition performed better on the number line estimation task at posttest than the participants with low scores in the nonlinear group ($t(11.832)=2.429, p=0.032$). Participants with high scores on the digit comparison task performed similarly on the number line estimation task regardless of if they were in the linear or nonlinear group (all $p>0.2$). Based on the results, the linear intervention was particularly useful for participants who started with low numeracy skills, giving further credence to the conclusion that the linear boxes scaffold is effective.

To further evaluate the individual differences on number line estimation error hierarchical linear models (HLMs) were constructed. The models took into account all factors that might affect changes in mean PAE (e.g. sex, grade-level). In addition to the ancillary factors that were considered, the models controlled for scores on the MANL pretest. When considering all potential predictor variables, including MANL pretest score, the average difference between the linear and nonlinear condition was no longer present at Phase 4, the posttest. This means once the predictors accounted for variation, the mean PAE for the linear and nonlinear group did not differ significantly.

A benefit of HLM is that it can provide information about the growth rate of a variable, in this case the mean PAE, in addition to giving a comparison of groups at a fixed time point. The ideal growth of mean PAE is to become more negative over time. The results for the main effects model demonstrated that the treatment group's (i.e. linear) mean PAE was decreasing at a faster rate (-0.09 logits PAE) than the nonlinear group ($p=0.005$). So while at posttest the linear and nonlinear groups did not differ, the difference in growth rates suggests that perhaps with more sessions they would have reached a point when they did differ significantly.

The MANL pretest scores being significant in the model led to further exploration. I used the median MANL pretest score to split the participants into two groups, low and high. The results showed that the low linear group's mean PAE was decreasing at a faster rate (-0.32 logits PAE, $p=0.007$) than the low nonlinear group's mean PAE. The high linear and nonlinear groups' growth rates did not significantly differ. The linear intervention was particularly useful for helping participants with low MANL pretest scores decrease their mean PAE at a faster rate.

Linearity as an Intervention

What factors besides feedback could contribute to the improvements in number line estimation ability? I propose that categorizing numbers into a straight line is an essential factor in the improvement in number line estimation error growth rate. Placing numbers into categories linearly reinforces key numerical properties. One numerical property that the linear boxes reinforce is the ordinal property. "…Numbers have an ordinal property in that they are sequenced" (Sarama & Clements, 2009, 85). Having participants identify numbers as really small, small, medium, big, and really big linearly encourages consideration of the sequence of numbers in a linear fashion and builds ordinality. Another numerical property that linear categorization reinforces is "later is greater" (Gunderson, Spaepen, & Levine, 2015): numbers that occur later in the count order or further to the right on the number line are greater in magnitude.

Another potential reason for the linear group's improvement in number line estimation skills is that the *representational mapping hypothesis* (Siegler & Ramani, 2009) extends beyond physical board games to computer-based number categorization tasks. The representational mapping hypothesis states that it is easier to learn the desired internal representation the more the physical representation transparently maps onto the internal representations (Siegler & Ramani, 2009). The mental number line is linear (Zorzi, Priftis, & Umiltà, 2002) and number line estimation is a visual representation of the internal representation of number (Elofsson, Gustafson, Samuelsson, & Träff, 2016): thus the linear categorization task led to improvements on number line estimation error. While the representational mapping hypothesis specifies the physical representations mapping to internal representations, this study demonstrates the connection between virtual (i.e. computer-based) representations and internal representations.

At posttest the linear group's error was less than the nonlinear group's in quintiles 1 and 2. Stated another way, the linear group outperformed the nonlinear group when estimating numbers that fell between 0 and 39 on the number line. One of the proposed representations of numerical development is a shift from a logarithmic or compressed number line to a more evenly spaced, linear number line (Siegler & Booth, 2004). In the shift from logarithmic to linear number representation smaller numbers improve first (Opfer & Siegler, 2007). Smaller numbers are the numbers that people encounter most frequently and with which people are most familiar. It follows logically that if an intervention on number is effective, it will be most effective first with the numbers with which children are most familiar. Larger numbers are more difficult thus participants need more time to learn the scaffold and apply it to the more challenging numbers.

Learning Study

As stated in chapter 1, this study was also a learning study (see Ginsburg, Labrecque, Carpenter, Pagar, 2015). What conclusions does the current study support about the activity's design? A significant finding is that this computer-based number categorization task is comparable to the categorization task done with baskets in Laski and Siegler's (2007) study. Further, the computer-based number line estimation task is also comparable to the paper and pencil (e.g. Elofsson, Gustafson, Samuelsson, & Träff, 2016) or alternative (e.g. Ebersbach et. al., 2008) number line estimation tasks that are often administered. When the number categorization task is coupled with the number line estimation task, number line estimation error can be reduced. However, the task's boxes should be arranged linearly for optimal performance on the number line estimation task. Two sessions with the linear intervention leads to changes in number line estimation error on the first 2 quintiles. Further study is needed to determine how many sessions are associated with changes in number line estimation error for all five quintiles.

Implications for the Classroom

Because magnitude knowledge is foundational to numerical development, teachers should incorporate tasks such as the number categorization and number line estimation tasks described here into their instruction. The tasks can act as an embedded assessment tool in that it gives a measure of students' magnitude understanding. Further, performance on the number line estimation task prior to categorizing numbers could help teachers identify students who are still using a logarithmic representation of number and for which numbers they are using that representation.

During early childhood education in school children begin to move from concrete representations of number (e.g. ten frames) to increasingly abstract number representations (e.g. traditional algorithms). Number line estimation could be a useful tool in transitioning from concrete to the more abstract in that it provides a visual representation of number and magnitude. Further, the task presented in this study could provide the first foray into proportional reasoning and rational number.

Chapter 5

Conclusion

The present study aimed to improve children's number line estimation skills through a linear number categorization intervention. Both the number line estimation task and the number categorization task were computer-based. Linear categorization was associated with improved number line estimation error, especially for the participants who started with low numeracy scores. This provides evidence that computer-based linear number categorization can be used to improve computer-based number line estimation. This is important because number line estimation is a foundational task which measures understanding of magnitude and the relationships between numbers. Further, number line estimation is associated with mathematics achievement. However, more study is needed to know if building number line estimation skills leads to higher scores on mathematics achievement measures.

The current study has limitations. The sample size of the current study made it difficult to find strong trends in the data. With a larger sample size it would have been easier to tease out possible group differences. Further, with a small sample size, generalizability of the results is difficult. Additionally, I administered the treatment for only two sessions and the participants in the linear group improved their estimation skills on the first two quintiles. An increase in treatment dosage may have led to growth on more quintiles.

In the future it would be beneficial to replicate this study with a larger sample and with the addition of a broad measure of mathematics achievement as a far transfer task. It would also be helpful to know how much participants used the midpoints of the boxes, the category prototypes, when categorizing to determine the utility of the demonstration videos that introduce the task. Additionally, the number of boxes that are used for the linear sort should be varied and tested. As Siegler and Opfer (2003) suggest, people may use different landmarks in different contexts. While quintiles were used here, quartiles, deciles (e.g. ten boxes) or some other configuration may work better. Finally, there is much to be learned about children's number line estimation strategies. It has been suggested that use of the origin and endpoints are some of the first estimation strategies children use (Ashcraft & Moore, 2012). However, from the growth happening in the first two quintiles and not the first and last quintiles, it suggests that participants were not using the endpoint to anchor their estimates. Future studies should incorporate a *clinical interview* (Ginsburg, 1981) to collect data on the types of strategies children are using to place their estimates.

References

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of experimental child psychology*, *111*(2), 246-267.

Bächtold, D., Baumüller, M., & Brugger, P. (1998). Stimulus–response compatibility in representational space. *Neuropsychologia*, *36*, 731–735.

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201-221.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation- evidence against a representational shift. *Developmental Science*, *14*(1), 125-135

Barth, H., Starr, A., & Sullivan, J. (2009). Children's mappings of large number words to numerosities. *Cognitive Development*, *24*(3), 248-264.

Booth, J. L. & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*, 189-201.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, 79(4), 1016-1031

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, 18(2), 331-338.

Cowan, R., & Powell, D. (2014). The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability. *Journal of Educational Psychology*, *106*(1), 214-229.

Creighan, S. (2014). Investigating the Effects of the MathemAntics Number Line Activity on Children's Number Sense. (Doctoral dissertation). Retrieved from https://clio.columbia.edu/dissertations

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371.

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.

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5-to 9-year old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology*, *99*(1), 1-17.

Elofsson, J., Gustafson, S., Samuelsson, J., & Träff, U. (2016). Playing number board games supports 5-year-old children's early mathematical development. *The Journal of Mathematical Behavior*, *43*, 134-147.

Epstein, D., Mendick, H., & Moreau, M. P. (2010). Imagining the mathematician: Young people talking about popular representations of maths. *Discourse: Studies in the Cultural Politics of Education*, *31*(1), 45-60.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307-314.

Geary, D. C. (2006). Development of mathematical understanding. *Handbook of Child Psychology*.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PloS One*, *8*(1), e54651.

Gelman, R., & Gallistel, C. (1978). The child's understanding of number. *Cambridge, Mass: Harvard University Press.*

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning disabilities*, *38*(4), 293-304.

Ginsburg, H. (1981). The clinical interview in psychological research on mathematical thinking: Aims, rationales, techniques. *For the learning of mathematics*, *1*(3), 4-11.

Ginsburg, H. P. (1997). Mathematics learning disabilities A view from developmental psychology. *Journal of Learning Disabilities*, *30*(1), 20-33.

Ginsburg, HP, Carpenter, K, & Labrecque, R (2011). Introduction to MathemAntics: Software for children from age 3 to grade 3

Ginsburg, H. P., Jamalian, A., & Creighan, S. (2013). Cognitive guidelines for the design and evaluation of early mathematics software: The example of MathemAntics. In *Reconceptualizing early mathematics learning* (pp. 83-120). Springer Netherlands.

Ginsburg, H. P., Labrecque, R., Carpenter, K., & Pagar, D. (2015). New Possibilities for Early Mathematics Education: Cognitive guidelines for designing high-quality software to promote young children's meaningful mathematics learning. In *The Oxford Handbook of Numerical Cognition*.

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: the role of the linear number line. *Developmental Psychology*, *48*(5), 1229-1241.

Gunderson, E. A., Spaepen, E., & Levine, S. C. (2015). Approximate number word knowledge before the cardinal principle. *Journal of experimental child psychology*, *130*, 35-55.

Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review*, *107*(3), 500.

Hubbard, E.M., Piazza, M., Pinel, P., Dehaene, S. Interactions between number and space in parietal cortex Nature Reviews. *Neuroscience*, *6* (2005), pp. 435–448.

Krasa, N., & Shunkwiler, S. (2009). *Number Sense and Number Nonsense: Understanding the Challenges of Learning Math*. Brookes Publishing Company. PO Box 10624, Baltimore, MD 21285.

Kuhn, D. (1995). Microgenetic study of change: What has it told us?. *Psychological Science*, 133-139.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, *78*(6), 1723-1743.

Luke, D. A. (2004). *Multilevel modeling* (Vol. 143). Sage.

Mix, K.S. & Cheng, Y.L. (2012). The relation between space and math: Developmental and educational implications. J.B. Benson (Ed.), Advances in child development and behavior, vol. 42, Academic Press, Burlington, VT, pp. 197–243

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519-1520.

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PloS One*, *8*(7), e67918.

Oakes, J. (1990). Opportunities, achievement, and choice: Women and minority students in science and mathematics. *Review of Research in Education*, *16*, 153-222

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, 55, 169-195.

Opfer, J. E., & Thompson, C. A. (2008). The trouble with transfer: Insights from microgenetic changes in the representation of numerical magnitude. *Child Development*, *79*(3), 788-804.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*(2), 375-394.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*(2p1), 274.

Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition*, *108*(1), 263-270.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. Routledge.

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*(2), 344-357.

Siegel, A. W., Goldsmith, L. T., & Madson, C. R. (1982). Skill in estimation problems of extent and numerosity. *Journal for Research in Mathematics Education*, 211-232.

Siegler, R. S. (2016). Magnitude knowledge: the common core of numerical development. *Developmental Science*, *19*(3), 341-361.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford University Press.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428-444.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, *46*(6), 606.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237-250.

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, *11*(5), 655-661.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*(3), 545.

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, *62*(4), 273-296.

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General*, *142*(1), 193.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89-96.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613-629.

Sullivan, J. L., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review*, *18*(3), 557-563.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80.

Umiltà, C., Priftis, K., & Zorzi, M. (2009). The spatial representation of numbers: evidence from neglect and pseudoneglect. *Experimental Brain Research*, *192*(3), 561-569.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*, 220–251.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1-B11.

Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *Journal of Cognition and Culture*, *5*(1), 165-190.

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Brain damage: neglect disrupts the mental number line. *Nature*, *417*(6885), 138-139.