Children's Number-line Estimation Deconstructed

by

Carley Graceanna Piatt

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology

University of Alberta

Abstract

Number-line estimation is an important, useful, everyday skill that has been linked to numerical cognition and mathematical achievement more generally. Despite numerous investigations in the last decade and the importance of number-line estimation as a mathematical concept, gaps remain in our knowledge of how number-line estimation develops. Averaging across individuals and ignoring trial-by-trial variability on the number-line estimation task may result in overlooking important information. I reviewed the number-line literature, identified several shortcomings in current knowledge, and developed and applied an alternative approach to study how 24 students in each of Grades 2, 4, and 6 make number-line estimates on 0-100 and 0-1000 lines. As in previous research, measures of accuracy and linearity on number-line estimation were taken as measures of implicit conceptual number-line knowledge. Explicit conceptual number-line knowledge was measured by assessing students' explicit understanding of, for example, proportions and scale on the 0-100 number line. To measure procedural knowledge self-reports were collected from students as they estimated targets on number lines on a tablet, and a task analysis of number-line estimation was used to guide the classification of number-line processes into solution procedures.

Several key findings emerged. First, children's explicit knowledge about the number line, such as their understanding of equal intervals and proportions, increased with age and was positively correlated with a linear pattern of number-line estimates. This result is important because it is the first time a measure of explicit conceptual number-line knowledge has been linked to performance on the number-line task with children in Grades 2, 4, and 6. Second, a task analysis of the processes used in number-line estimation guided the identification of how processes were combined into solution procedures. The task analysis allowed for the coding of

observations and students' self-reports, which revealed immense variability in the solution procedures students used.

Further, identifying students' solution procedures paved the way for successfully identifying *tactics*, the profile of solution procedures students selected as a function of target. As expected, older students used more conceptually advanced tactics compared with younger students. Even when controlling for grade, the use of advanced tactics was generally related to having more explicit number-line knowledge. Moreover, distinct patterns of discrepancy in estimation emerged as a function of tactics, giving rise to the conclusion that number-line estimation is a product of not only implicit conceptual number-line knowledge but also what children explicitly know about the number line and the procedures they used to estimate.

Finally, having mapped the ways in which students estimated, I investigated the ways in which students adjusted their number-line estimation tactics across two ranges and found that successful adjustment from the 0-100 to the 0-1000 line was often associated with more explicit conceptual number-line knowledge. Moreover, I found that several kinds of shifts in tactic led to adaptive performance on the number-line task. Overall, these results shed new light on what develops in children's number-line estimation by creating a window into children's explicit conceptual and procedural number-line knowledge.

Preface

This research is an original work by Carley Piatt. The research project of which this dissertation is a part, received ethics approval from the University of Alberta Research Ethics Board, Project Name "Reasoning About Estimation in Children," No. 00024310, 21 October 2011.

Dedication

For Gage

On the contrary, we succeed in thinking well by thinking hard; we get the valuable thought-variations by concentrating attention upon the body of related knowledge which we already have; we discover new relations among the data of experience by running over and over the links and couplings of the appreciative systems with which our minds are already filled; and our best preparation for effective progress in this line or in that comes by occupying our minds with all the riches of the world's information just upon the specific topics of our interest.

~ J. M. Baldwin in On Selective Thinking,1898

Acknowledgements

I must first thank Drs. Jeffrey Bisanz and Joanne Volden for agreeing to supervise me and Drs. Elena Nicoladis and Lonnie Zwaigenbaum for agreeing to be on my committee. You all said “yes,” sight unseen, and I am so grateful for your belief in me at the start, and every step of the way, as I charted my way forward as a “cognitive cartographer.” I can imagine no better mentor than Jeff for this journey and am grateful for his teaching me to think productively and subtly about children’s thinking. Both Jeff’s and Joanne’s keen insights and unique perspectives made the work richer and more nuanced than I could have ever imagined or accomplished on my own. Joanne’s belief in the value of this work for understanding cognitive development in ASD propelled me forward more times than I can count. Thank you, Jeff and Joanne. Elena and Lonnie brought creative and constructive comments to every meeting and their passion for their respective fields made them ideal mentors. I would also like to thank Drs. Patrick Lemaire and Sandra Wiebe for being my external examiners and for their thoughtful and thought-provoking questions.

A sincere thank you to Michael Choi for designing and debugging the number-line app that was used in this research. Without his willingness to work on this project, the road forward would have been much longer and less innovative. Deeply felt thanks are in order for Marian Coret, who spent many hours collecting, coding, and managing data for this project. Truly, without her practical help and her constant encouragement and reminders to have fun, drink more coffee, and celebrate the small accomplishments, this project wouldn’t be half of what it has turned out to be. Thanks also to Megan Cook and Taylor Fehr who helped code the number-line knowledge data (the superhero measure!). Thank you to Jody Wolfe and Sue Piatt for reading a draft of the dissertation and for giving me helpful edits and feedback. This research was supported by funding from the Women and Children's Research Institute, Canadian Institute of Health Research, and from several awards from the Department of Psychology. Finally, without the willingness of the students, teachers and staff in several schools in the greater Edmonton area, there would be no ideas to fill these pages.

In the Laboratory for Research on Mathematical and Cognitive Development, I am indebted to my past lab mates, Rebecca Watchorn Kong and Anna Matejko. Becky's energy for life, tenacity for research, commitment to friendship and citizenship, and insights into the grad program grounded me and guided me toward the light at the end of the tunnel in more ways and more often than she knows. Anna's insights and questions both about how children learn math and about living life fully make her a valued colleague that I am lucky enough to also call a good friend. A special thanks to Mary Spytz for her work on the exploratory study of math in ASD and for her friendship and inspiring passion for exploring the intersection between art and math. I would also like to thank several colleagues in the department: all of the gentleman in the shop, Isaac, Lou, Al, and Philip, who made sure the machines worked, kept me in paper, and always greeted me (and any question or request) with kindness and competence, and the ladies in the general office, past and present, who make the department tick and make long days a little brighter. To my friends and colleagues in the department who held me accountable for my writing goals in the last few years, thank you for rewarding my efforts with coffee and your good company.

I am deeply indebted to my friends and family near and far who have brought balance and beauty to my world: Sandra, Aunt Louie, Kala, Daina, Svenja, Robin, the YKers, the Cheema family, and Jody and Shane who supplied encouragement, space, treats, and the company of sweet bunnies for many long days and nights of writing. And lastly, without the love and support of my parents and husband, the frustrating times would have been unbearable and the moments of celebration would have meant nothing. To my parents, Sue and Terry, thank you for laying the foundation of love and humor and for nurturing me to be confident, creative, and curious in equal measures, without which I could not have done this work. To my husband and biggest cheerleader for the last five years, Sim, thank you for the love, support, encouragement, patience, cooking dinners, hugging out the crazy, making me laugh, and being my rock of positivity. Your unfailing belief in me illuminated my path forward every day. Thank you.

Table of Contents

List of Tables

List of Figures

Children's Number-line Estimation

Estimating is an important, useful, everyday skill (Siegler & Booth, 2004; LeFevre, Greenham, & Waheed, 1993; Newman & Berger, 1984; Opfer & Thompson, 2008; Siegler & Booth, 2005). Estimation is the translation "between alternative quantitative representations, at least one of which is inexact" (Siegler & Booth, 2005, p. 198). Number-line estimation is the estimation of the physical position on a number line corresponding to a given number, or vice versa. In the last decade number-line estimation has attracted intense research interest for three reasons. First, the number line is an important tool for teaching students about arithmetical concepts including natural and real numbers (Heefer, 2011). Second, researchers have found that performing well on number-line estimation is associated with doing well on other kinds of mathematical tasks (Booth & Siegler, 2006; Laski & Siegler, 2007), with learning arithmetic (Booth & Siegler, 2008), and with obtaining higher mathematics achievement scores (Ashcraft & Moore, 2012; Booth & Siegler, 2006, 2008; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013; Schneider, Grabner, & Paetsch, 2009; Siegler & Booth, 2004; Träff, 2013). As a result, the number line is thought to be a concept central to organizing numerical knowledge (Booth & Siegler, 2008; Siegler, Thompson, & Opfer, 2009). Third, studying how students make their number-line estimates may reveal important insights about children's cognitive development because estimation evokes mathematical inventiveness as students must go beyond routinely using procedures and flexibly apply their mathematical knowledge (Newman & Berger, 1984; Siegler & Booth, 2005). By understanding how children estimate on number lines, it may be possible to look, in detail, at the mathematical inventiveness afforded by the task and characterize the extent to which children think flexibly on the task.

I developed and applied an alternative approach to study how students in Grades 2, 4, and 6 make number-line estimates on 0-100 and 0-1000 lines. Several key findings emerged. First, children's explicit knowledge about the number line, such as their understanding of equal intervals, scale, and proportions, increased with age and was positively related to accuracy on the number-line task. This result is important because it is the first time a measure of explicit conceptual number-line knowledge has been linked to performance on the number-line task with children in Grades 2, 4, and 6.

Second, a task analysis of the processes used in number-line estimation guided the identification of how processes were combined into solution procedures. The task analysis allowed for the coding of observations and students' self-reports, which revealed both that students rely on procedural knowledge to estimate and that there was immense variability in the solution procedures students used. Further, identifying students' solution procedures within a number-line range paved the way for identifying tactics, the profile of solution procedures students selected as a function of target. As expected, older students used more conceptually advanced tactics compared with younger students. The use of advanced tactics was generally related to more explicit number-line knowledge, even when controlling for grade. Number-line estimation tactics were organized to reflect a tractable development sequence and results from this work generally aligned with the proposed sequence. Moreover, distinct patterns of estimation emerged as a function of tactics, leading me to conclude that number-line estimation may be a product of not only implicit conceptual knowledge but also procedural knowledge.

Third, even when controlling for grade, a general measure of intelligence, and measures of implicit conceptual number-line knowledge, my measure of explicit conceptual number-line knowledge was positively correlated with a measure of math achievement. As a result I suggest that the kinds of number-line knowledge that generalize to other kinds of numerical knowledge are not captured exclusively by measuring accuracy or linearity on number-line estimation. Finally, having mapped the ways in which students estimated, I investigated the ways in which students adjusted their number-line estimation tactics across two ranges and found that successful adjustment from the 0-100 to the 0-1000 line was associated with more explicit conceptual number-line knowledge. Moreover, I found that several kinds of shifts in tactic, or strategies, led to adaptive performance on the number-line task. Overall, these results shed new light on what develops in children's number-line estimation by creating a window into children's explicit conceptual and procedural number-line knowledge.

Number-line Estimation

Origins of the number line. The number line is primarily a Western cultural construction (Carey, 2001, in Heeffer, 2011) usually thought of as a horizontal line with higher numbers on the right. Heeffer (2011) defines a number line as "a *representation* of numbers on a straight line where points represent integers or real numbers and the distance between the points match the arithmetical difference between the corresponding numbers" (p. 2). The number line may have first appeared in print in John Wallis' (1685) work on algebra (Heeffer, 2011). Heeffer (2011) notes that, despite having the potential to be useful as "a model for reasoning, teaching, and understanding concepts" (p. 3), the number line was only introduced in mathematics education in the 1950s.

The most frequently used method for assessing children's understanding of number lines has been to evaluate the nature of their estimation patterns, usually by fitting mathematical models (e.g., Ashcraft & Moore, 2012; Barth & Paladino, 2011; Booth & Siegler, 2006; Bouwmeester & Verkoeijen, 2012; Opfer & Siegler, 2007; Rouder & Geary, 2014; Siegler &

Opfer, 2003; Siegler & Booth, 2004; Slusser, Santiago, & Barth, 2013; Young & Opfer, 2011). For example, if a child "understood" a 0-10 number line as defined by Heeffer, she should estimate the target on the line accurately, as shown in the top half of Figure 1. On a graph, with her estimates shown on the y-axis and the actual number on the x-axis, her estimates will align perfectly with the actual number and the resulting plot of her data will be highly linear (perfectly linear in this example). Thus, her understanding of this number line would be termed "linear."

For infants space and number are linked such that as one dimension increases, so does the other (de Hevia, Izard, Coubart, Spelke, & Streri, 2014). Infants, however, do not have a perfectly linear representation of numbers. Dehaene, Izard, Spelke, and Pica (2008) suggest that, for humans, numbers map onto space but cultural experiences alter that mapping. Infants show evidence of thinking of small numbers as being quite distinct from one another and of larger numbers as being less distinct and lumped together (for a review see Noël, Rousselle, & Mussolin, 2005). Descoudres (1921) described how preschool children can differentiate the numbers one, two, and three in a variety of tasks, but that around the number four they become less reliable and refer to numbers larger than four as "a lot." She called this pattern the *un, deux, trois, beaucoup* phenomenon (as cited in Gelman & Gallistel, 1978). As a result of thinking of small numbers as distinct from one another and larger numbers as lumped together, a young child with an internal, immature 0-10 number-line representation, may estimate the target on the line at incorrect spots that are more dispersed for smaller numbers and compressed for higher numbers, as shown in the bottom half of Figure 1. When these estimates are graphed, her number-line representation will appear *curvilinear*, better fit by a logarithmic than a linear equation. The nature of the errors will tend to be systematic and fit "Weber's law, a ubiquitous psychophysical law whereby increasingly larger quantities are represented with proportionally greater imprecision, compatible with a logarithmic internal representation with fixed noise" (Dehaene et al., 2008, p. 1217).

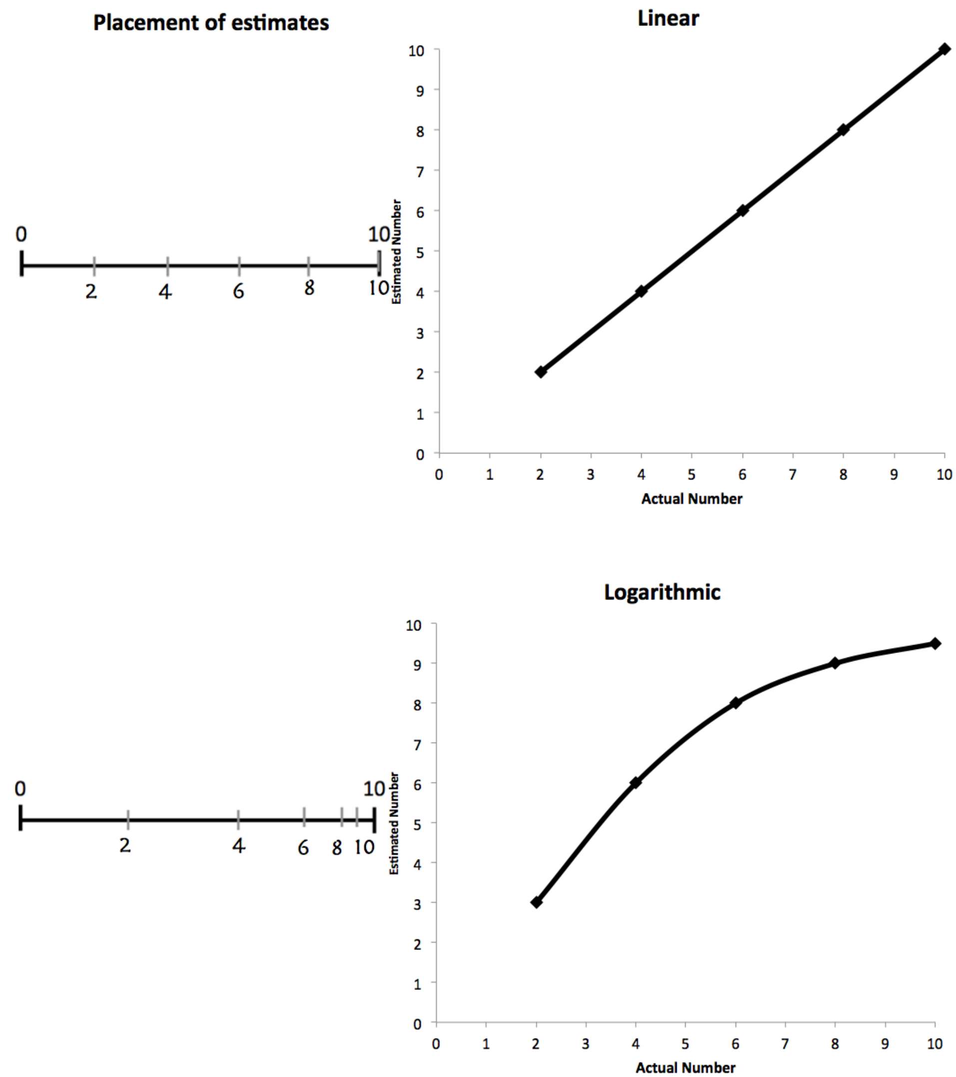


Figure 1. Examples of linear (top half) and logarithmic (bottom half) estimates on 0-10 number line with illustrative estimates on the 0-10 line shown on the left and patterns of estimates as a function of target shown on the right.

Dehaene et al. (2008) investigated the role of culture on number-line representation with the Mundurucu, an Amazonian indigene culture. The Mundurucu have little formal education, limited number language, and no access to rulers, measurement devices or graphs. Still, they have a sophisticated nonverbal concept of space and number as shown by, for example, their ability to reason about ideas in Euclidean geometry (Izard, Pica, Spelke, & Dehaene, 2011). To investigate their understanding of number-line estimation, 33 Mundurucu people, including 22 adults, pointed to the positions of numbers, in four different modalities (dots, tones, Mundurucu and Portuguese numerals) on a 0-10 line. The Mundurucu's estimation patterns were more *curvilinear* as compared with the linear estimation patterns of 16 American adults (Figure 2). Though the Mundurucu's pattern of estimates may be viewed as immature in a North American context, the pattern is more likely due to differences in specific mathematical experiences, such as measuring with rulers, than to immature development of mathematical cognition, given the Mundurucu's sophistication in reasoning about geometry.

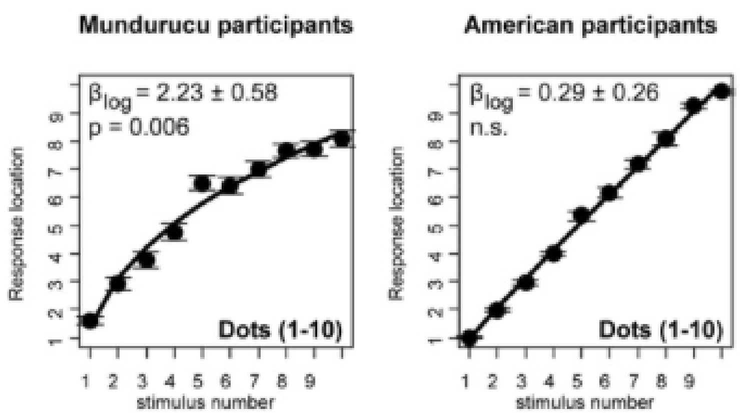


Figure 2. Comparison of Mudurucu and American participants estimating the location of targets (presented as a group of dots) on 0-10 line. Reprinted portion of Figure 2 from "Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures" by S. Dehaene, V. Izard, E. Spelke, and P. Pica, 2008, *Science, 320*, p.1219. Copyright 2008 by the American Association for the Advancement of Science.

When 4-year-old Italian children estimated on a 0-10 line, their estimates were more curvilinear and better fit by a logarithmic than a linear model (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010). In contrast to the Munducuru, by the time Italian children are about five years old their estimates of 0-10 are more accurate and a linear model fits nearly perfectly (Berteletti et al., 2010). This points to the potentially strong and early effect that education and enculturation have on children's representation of a number line, at least for the 0-10 line. For other number-line ranges, such as 0-100 and 0-20, the kindergarteners showed a curvilinear pattern of estimates, best fit by a logarithmic model (Berteletti et al., 2010). These results are consistent with numerous studies in the last decade showing that children's patterns of number-line estimates change as a function of the number-line range and the child's grade (Booth & Siegler, 2006; Opfer & Siegler, 2007; Rouder & Geary, 2014; Siegler & Opfer, 2003; Siegler, Thompson, & Opfer, 2009). Such evidence has been used to point, in psychological terms, to an abrupt and distinct shift in representing number, called the logarithmic-to-linear shift, (Dehaene et al., 2008; Siegler & Opfer, 2003). Further to the argument that educational experiences play a role in facilitating the logarithmic-to-linear shift is evidence demonstrating that the shift occurs between kindergarten and Grade 2 for the 0-100 scale, between Grades 2 and 4 for the 0-1000 scale, and between Grades 3 and 6 for the 0-100,000 scale (for a review see Siegler, Thompson, & Opfer, 2009). Evidence, however, is mounting that calls into question an abrupt logarithmic-to-linear shift as the best model for the development of number-line estimation.

Methods for studying number-line estimation. Researchers measure number-line estimation in two ways (Siegler & Opfer, 2003). In both methods participants are shown a number line with marked endpoints. In the number-to-position (NP) task, participants are shown a target number and asked to estimate the target number's position on a number line. In the position-to-number (PN) task, a target position is already marked on the line and participants are asked to estimate the number that corresponds to the target position (Newman & Berger, 1984;

Pettito, 1990; Siegler & Opfer, 2003). Of the two, the number-to-position (NP) task has been used more frequently (Booth & Siegler, 2005; Young & Opfer, 2011). Skill in number-line estimation is often captured in two ways.

First, to assess how well a student estimates each target number within a range, the percent absolute error or PAE (Booth & Siegler, 2006, 2008; Opfer, n.d.; Petitto, 1990) is calculated:

$$PAE = \frac{|\text{Estimated Position} - \text{Target Presented}|}{\text{Numerical Range}} \times 100$$

For example, if a student, estimating the location of 42 on a 0-100 number line, placed the mark at the location that corresponded to 55, his percent absolute error on that trial would be |42-55|/100 multiplied by100, equalling 13%. Second, to capture how linear a student's estimates are within a range, her estimates for each target are plotted as a function of the number's actual position, where perfect accuracy on the task would be fit by a linear model with a slope of 1.00 and $y = x$ (Young & Opfer, 2011). The plots can be assessed for how well different kinds of mathematical models, such as linear or logarithmic equations, fit the data. Investigations of students' estimates on the number-line task consistently show: (a) children's error decreases with age, especially on number-line ranges with which they may have more experience or exposure (Siegler & Ramani, 2009; Opfer & Thompson, 2008), (b) on average, younger children show a more curvilinear pattern of estimates for larger ranges (e.g., 0-1000) compared with older children, whose pattern of estimates appear to be more linear (e.g., Ashcraft & Moore, 2012; Barth & Paladino, 2011; Berteletti et al., 2010; Siegler & Opfer, 2003; Siegler, Thompson, & Opfer, 2009; Slusser et al., 2013), and (c) when faced with a larger, less familiar line such as 0-100,000, Grade 3 students show a pattern of estimation that is more curvilinear than linear

(Thompson & Opfer, 2010). Four models, all based on fitting equations to patterns of estimates, have been proposed.

Psychological models of number-line estimation. As noted, the prominent psychological model used to explain the pattern of change in the data is called the logarithmic-to-linear shift. Siegler, Opfer, and colleagues (Booth & Siegler, 2006; Siegler & Opfer, 2003; Siegler et al., 2009; Young & Opfer, 2011) argue that students recruit one of two distinct representations of numerical magnitude when estimating targets on a number line. The less mature representation occurs when students represent smaller numbers as being more distinct, with more space between numbers, whereas larger numbers are represented less precisely and as more compressed on the line (Dehaene et al., 2008; Siegler & Opfer, 2003; Siegler, et al., 2009). When students use this immature representation, the pattern of estimates appears to increase logarithmically ($y = k \times \ln x$) as shown in Figure 1. The logarithmic function is just one way of capturing number-line compression at higher numbers; in some studies, other power functions (e.g., Ashcraft & Moore, 2012; Bouwmeester & Verkoeijen, 2012) provide a better fit to the data than logarithmic functions (Rouder & Geary, 2014). In contrast, a more mature representation emerges when children presumably learn more about the number line and come to understand that numbers are separated by equal intervals across a line. As a result, the pattern of children's estimates increases linearly with actual magnitude ($y = x + b$) (Siegler et al., 2003, 2004, 2006; Young & Opfer, 2011). Over time and with experience, children seem to replace the immature underlying representation that manifests as a curvilinear line with a more mature, linear underlying representation—the so-called logarithmic-to-linear shift (Siegler & Opfer, 2003; Siegler et al., 2009).

Siegler and colleagues, and others, have found evidence for this apparent logarithmic-to-linear shift across different ages, scales or ranges, and kinds of estimation tasks (e.g., Berteletti, et al., 2010; Booth & Siegler, 2006; Siegler & Opfer, 2003; Siegler et al., 2009). For example, students in kindergarten are less accurate in their estimates of targets on a 0-100 number line than Grade 1 students (Booth & Siegler, 2006). In kindergarten, children's estimates are better fit by a logarithmic than a linear function on 0-100. In contrast, Grade 1 students show a pattern of estimates better fit by a linear than a logarithmic function on 0-100 (Booth & Siegler, 2006). This pattern of younger students relying on logarithmic representations for unfamiliar ranges has been reported to shift in such a way that even older students, such as Grade 3 students, will recruit a less mature, logarithmic representation when faced with an unfamiliar range such as 0-100,000 (Siegler et al., 2009).

At least three strengths are associated with the logarithmic-to-linear shift model. First, the model of a discrete shift from a less to a more mature representation has been argued to be rooted in, and is consistent with, biologically based accounts of a non-symbolic Approximate Number System mechanism that enables even young infants to discriminate among sets of objects with sufficiently different ratios (Siegler & Lortie-Forgues, 2014). Second, many behavioral studies of number-line estimation have reported results consistent with the logarithmic-to-linear shift (Booth & Siegler, 2006; Siegler & Booth, 2004; Siegler & Opfer, 2003; Young & Opfer, 2011). Finally, there are several studies illustrating that familiarity and learning about important concepts associated with number lines such as counting (Newman & Berger, 1984; Siegler & Ramani, 2009) and numerical categorization (Laski & Siegler, 2007), are associated with a logarithmic-to-linear shift. Despite the strengths of the logarithmic-to-linear shift account, shortcomings of the model have been identified and addressed with other kinds of model-fitting approaches.

One proposal is that the logarithmic pattern seen in young children's data is better described in terms of two linear functions with differing slopes (Young & Opfer, 2011). Instead of a shift-in-representation account, differences in children's thinking about the number line may be the result of differences in familiarity with certain numbers, counting ability within a range (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008), or whether the numbers have one or two digits (Moeller, Pixner, Kaufmann, & Nuerk, 2009). That is, different sections of the number line may be best fit by several linear models with differing slopes (Ebersbach, et al., 2008; Moeller et al., 2009). For example, Ebersbach et al. suggested two linear segments, one with a steep slope to capture a student's familiar numbers and another with a flatter slope to capture his unfamiliar numbers. Some evidence for the segmented model account has been found (Ebersbach et al., 2008; Moeller et al., 2009) but segmented models are difficult to distinguish from logarithmic models (for a comparative analyses see Young & Opfer, 2011). Moreover, the appeal to several models to account for patterns of estimates within the same range has been criticized for being unnecessarily complex (Barth & Paladino, 2011; Young & Opfer, 2011).

Citing model complexity as a drawback in both the segmented linear and logarithmic-to-linear accounts, Barth and colleagues (Barth & Paladino, 2011; Slusser et al., 2013; Sullivan, Juhasz, Slattery, & Barth, 2011) suggested that a single model of proportional reasoning, rooted in psychophysics, may better characterize children's changing patterns of accuracy. Barth and Paladino (2011) argue that (a) evidence for the logarithmic-to-linear shift is not always found in estimation tasks without number lines, such as when young children in preschool and kindergarten estimated the number of dots on a card (Barth, Starr, & Sullivan, 2009; Lipton & Spelke, 2005), (b) the seemingly "linear" responses of older students often show patterns of over- and underestimation not explained by a logarithmic-to-linear shift, and (c) number-line estimation tasks require judgments about proportions. Barth and colleagues specifically reason that as children develop the ability to judge proportions, instead of estimating a point on a line, they estimate a part of a whole (Barth & Paladino, 2011; Slusser et al., 2013; Sullivan et al., 2011). Drawing from and extending the work of researchers studying proportional thinking in other domains (e.g., Spence, 1990; Hollands & Dyre, 2000), Barth et al. suggested that if students make a proportional judgment when estimating, then their estimation patterns should look similar to the patterns of proportional judgment observed on other tasks.

In this line of reasoning, if students are not making a proportional judgment, a single power function, such as the logarithmic function, can account for the pattern of compressed estimates that emerges as students rely on their internal, but immature, sense of the number line. As students experience and learn about numbers across larger ranges, such as 0-100, they learn to estimate using proportional judgments (Barth & Paladino, 2011). To do so, they must estimate the relation between two distances, the distance from 0 to the end of the line, and the distance from, for example, 0 to the target. When the student combines the two estimations, the whole line compared with the part of the line from one end to the target, a proportional judgment results. The resulting pattern of estimates is not a logarithmic function, corresponding to a compressed model, but an s-shaped function that is accurate at zero, the end point, and the midpoint with overestimations between zero and the midpoint and underestimations from the midpoint to the endpoint. This model has been called the *one-cycle* or *two-anchor* model because students presumably anchor to the two ends and show one cycle of over- and under-estimation as a result.

In other words, as children become increasingly knowledgeable about the number line, they may begin to use an increasing number of reference points or anchors (Barth & Paladino, 2011; Rouder & Geary, 2014). Initially children may anchor to just the low end of the number line, resulting in a compressed pattern of estimates (Rouder & Geary, 2014). Rouder and Geary refer to this as the *one-anchor* model. Later children may anchor to both ends of the number line, sometimes called the *two-anchor* model, resulting in more accurate estimates near the ends and around the middle as using the ends may implicitly give rise to a tendency to visually portion the line into two symmetrical halves (Barth & Paladino, 2011; Rouder & Geary, 2014). Next, children may come to explicitly use the midpoint as they estimate, resulting in being most accurate near the reference points. This model is called the *two-cycle* or *three-anchor* model because students may explicitly anchor to the ends and midpoint (Rouder & Geary, 2014), and as a result may implicitly be relying on the midpoints between 0 and half (quarter) and between half and the high end (three quarters). As a result, these students would show two cycles of over- and under-estimation. Arguably, as children begin to use more anchors, the best-fitting equation becomes increasingly linear (for reviews see Barth & Paladino, 2011; Rouder & Geary, 2014).

The appeal of the proportional-reasoning model to explain the development of number-line estimation is that the mental number line used by children is invariant. What changes is the number of reference points children use. If this view is correct, the age-related increase in linearity does not reflect a different, linear representation but rather is an epiphenomenon resulting from students' learning to use more reference points. Several researchers have found evidence for the proportional-reasoning model by looking at patterns of accuracy on the number-line task (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Bouwmeester & Verkoeijen, 2012; Rouder & Geary, 2014; Sullivan et al., 2011), especially when considering individual students' data (Rouder & Geary, 2014; Slusser et al., 2013).

Bouwmeester and Verkoeijen (2012) highlighted five methodological issues with the logarithmic-to-linear model fitting approach and introduced a latent variable modeling approach as an alternative method for fitting models to estimation patterns. Three of the five methodological issues they raised are derivatives of the criticism that important information can be missed when averaging across data (Siegler, 1987; Simon, 1975). For example, they noted that in averaging across individuals, variability is assumed to be error rather than "important information about the estimation patterns of the individuals" (Bouwmeester & Verkoeijen, 2012, p. 249), and that collapsing within age groups assumes that all children of the same age are developmentally similar, again treating variation as unsystematic error. These concerns have been shown to be legitimate in other areas of mathematical thinking where a close look at individual differences revealed a more detailed and accurate story about how arithmetic skill develops (e.g., Siegler, 1987).

To capture some of the individual variability using a data-driven approach, Bouwmeester and Verkoeijen (2012) used latent class regression analysis to identify classes of students who showed the same estimation patterns on a 0-100 number line. With data from students in kindergarten through Grade 2, they found a solution with five models captured 83% of the variance. Of the five models, a highly linear model accounted for a quarter of the students (whose mean age was 7.64 years). A model that looked similar to the three-anchor proportional-reasoning model, with high accuracy at the ends and middle, over-estimation on the low end and underestimation on the high end, accounted for a third of students (with a mean age of 7.63 years). Two more models each accounted for about a fifth of students. For one group (mean 7.41 years), the model appeared quite linear but was best fit by a cubic function with little error across the line except some underestimation for smaller numbers. For a slightly younger group (mean

7.03 years) the function looked fairly similar to the logarithmic group described elsewhere in the literature but was better fit by a cubic function. The fifth model, a horizontal linear model indicating that students estimated around the middle across the line, applied to just three young students (mean 6.4 years).

Evidence for multiple models of estimation patterns, several of which do not correspond to and cannot be accounted for by the logarithmic-to-linear shift, raises the question, "If estimation patterns are not *just* a function of internal representation, *what else* do they reflect?" Bouwmeester and Verkoeijen (2012) suggested, for example, that the number of reference points students used may influence estimation patterns but did not test this directly. Using data from a longitudinal study with students in Grades 1 through 5 who estimated on a 0-100 line, Rouder and Geary (2014) fit three models, one- and two- and three-anchor models, to individual students' accuracy data. In Grade 1, 63% of students had patterns of estimates best fit by the two-anchor proportional model. By Grade 2, plurality of children had patterns of estimates best fit by the three-anchor model, and by Grade 5 the three-anchor model was the best fit for 58% of the children's estimation patterns. These results further suggest that how children are approaching the task, at least in terms of the kinds of reference points students might use, may reflect the observed patterns of estimates. Some researchers have considered other measures such as students' solution procedures on a trial-by-trial basis within a range to capture *what else* may contribute to the individual patterns of estimation observed at different ages, for different ranges.

Alternative methods for studying number-line estimation. Two of the earliest studies of number-line estimation focused on how students made their estimates. After having students estimate the position of an arrow on a 1-23 vertical number line, Newman and Berger (1984) asked students in kindergarten and in Grades 1 and 3 to explain how they made their estimates on a small, medium, and large target on the same line. Newman and Berger identified four types of estimation: (a) guessing, (b) counting up from the bottom-end point, (c) counting down from the top-end point, and (d) counting from the middle. Next they classified the pattern of each student's use of the four estimation types to characterize the student's approach to estimating across the three targets: the small, medium and large numbers. Four estimation rules emerged: (a) no reliance on counting (Rule 1), (b) counting forward only (Rule 2), (c) flexible use of both forward and backward counting (Rule 3), and (d) Rule 3 plus the use of reference points when endpoints did not seem helpful (Rule 4). Newman and Berger ordered the estimation rules from less (Rule 2) to more (Rule 4) sophisticated based on how well students' rules minimized counting, such that only counting forward for all targets (Rule 2) was less sophisticated than flexibly using different reference points and counting either up or down for the three targets (Rule 4). Older students reported using more sophisticated estimation rules such as Rule 4 more often than younger students. Even when controlling for age, students using a more sophisticated, flexible estimation rule had greater accuracy.

Newman and Berger's (1984) work illustrated two important points. First, there was variability in how individual children, within the same grade, estimated targets across the number line. Second, the amount of error in estimation was related to how flexibly students were thinking. Integrating the results of this study into the body of recent number-line literature is difficult because Newman and Berger used a vertically oriented line of a restricted range (1-23) presented on a computer, whereas many recent studies have used horizontal number lines of a variety of ranges (0-10, 0-100, 0-1000, and 0-10,000) with paper and pencil. Moreover, Newman and Berger limited their investigation to early-elementary students' self-reports on just three targets. Nonetheless, this study serves as a foundational study of children's number-line estimation.

In a second important study, Petitto (1990) had students in Grades 1-3 estimate three targets on each of two horizontal number lines: 0-10 and 0-100. Petitto observed what students were doing as they estimated, and they recorded students' accuracy on the task. Petitto identified four types of responses: (a) counting from left to right, (b) counting from right to left, (c) using the midpoint position, and (d) no visible strategy used. Like Newman and Berger (1984), Petitto evaluated each student's responses across a range, and classified each student into one of three strategy groups: unidirectional counting, use of midpoint, and interval counting (in the case of 0-100). Of note, when Petitto classified a student as using unidirectional counting, she also assessed whether students' use of unidirectional counting was appropriate "based on the relationship between the direction of counting and the position of the target" (p. 67). For example, a child was classified as inappropriately counting if he counted from the left endpoint to the rightmost target or from the right endpoint to the leftmost target. Petitto noted that inappropriate counting reflects thinking that is inflexible and not sensitive to the context of the task.

Petitto's (1990) results reflect those of Newman and Berger (1984). On the 0-100 number-line older students used the midpoint strategy more than younger students. Younger students used inappropriate unidirectional counting almost twice as often as older students, but all students used appropriate unidirectional counting equally. Moreover, older students adjusted their counting by using intervals of tens (compared with counting by ones) more than younger students. Petitto concluded that while all students counted as part of making their estimates, older students were more flexible in their ability to adjust the direction and interval they used to count.

Overall, all students had lower mean PAEs for the 0-10 than the 0-100 line but there were differences in accuracy as a function of target position. On 0-100, younger students had higher deviation scores for the two middle and high-end targets than older students. Petitto argued that this increased error on the upper half of the line was consistent with the idea that students initially used counting strategies from the 0 end of the line, resulting in greater error farther from 0. She went on to suggest that when students began to use proportions, the errors for the middle and high-end targets would be reduced. The relation between Petitto's strategy classifications and students' accuracy was not reported.

More recently, Schneider and colleagues (2008) tracked the eye movements of students in Grades 1 to 3 as they estimated on a 0-100 number line. They found that students looked more at the end points and middle of the line, suggesting that students used these areas of the line as part of their estimating processes. Looking at a point on the line is taken as an index of an orienting process (Schneider et al., 2008). Schneider et al. (2008) found that the precision of orienting increased with age such that younger children fixated to the ends and middle more whereas older children also fixated to other segments of the line between the ends and middle. This method provided behavioral evidence that students used different orienting points along the line, thus confirming that students may use landmarks in estimation (Petitto, 1990; Siegler & Opfer, 2003). Missing from this study is information about whether strategy choice varied on a trial-by-trial basis with, for example, information from self-reports (Schneider et al., 2008), as well as links to how strategy selection on specific targets relates to overall estimation patterns within a number-line range.

White and Szűcs (2012) moved toward describing children's number-line estimation strategies by examining how students in Years 1-3 (mean ages 6.4 to 8.5 years) estimated 8 targets on a 0-20 number line, presumed familiar to all students. White and Szűcs used analysis of variance to investigate whether students' estimation error varied as a function of Year (1, 2, or 3) and Target (2, 4, 7, 8, 11, 13, 16, or 17). As expected, they found that as year or age increased, estimation error decreased, but not significantly. What was novel, however, was a main effect of *target* on estimation error for all years. From their post-hoc analyses of individual targets White and Szűcs concluded that two targets, 11 and 13, specifically "might be the target of selective strategy use" (p. 7) because older students (Years 2 and 3) estimated these targets more accurately and with less variability than younger students (Year 1).

White and Szűcs (2012) did not fully explain the nature of the different strategies students might select that lead to the observed differences in estimation error, but they suggested that an in-depth analysis of how strategies are applied to targets would be fruitful. They also suggested how and when students anchor to different points on the line may change. Younger students may begin by anchoring to the 0 end, followed by anchoring to the 0 and high end, and eventually partition the line to generate other anchors such as the middle. Their results illustrate the need to consider, in light of a target's proximity to external or mental anchoring points, the unique behaviors different targets may afford. Notably, White and Szűcs also pursued a model-fitting approach and examined the fit of logarithmic, linear, and power models but ultimately questioned whether it is meaningful to focus on modeling the mental representation of the number line when there is evidence that other factors such as familiarity with a range, target location, knowledge of arithmetic, ability to partition the line and individual differences may also play a role in estimation.

Ashcraft and Moore (2012) employed a similar approach to that of White and Szűcs (2012). Ashcraft and Moore examined the estimation errors of students in Grades 1-5 and college as a function of 26 targets on a 0-100 number line. Consistent with previous literature, they found that error decreased with grade but also that estimation error changed as a function of target depending on the students' grade (even when excluding the college students). Ashcraft and Moore averaged each child's two estimations closest to 0, 25, 50, 75 and 100 (which they labeled origin, first quartile, midpoint, third quartile, and endpoint, respectively) to investigate the overall contours in each child's pattern of estimates. In so doing, they found that Grade 1 students showed low error near the origin, increasing error across the line, and slightly less error for targets near the endpoint. Grade 2 students showed low error at both endpoints. Grade 3 students showed low error at both endpoints and at the midpoint, creating an "M-shaped" pattern. Grade 4 students showed the same M-shaped pattern of error along with an attenuation of error at the first and third quartiles. This M-shaped pattern was observed in Grade 5 and college students too, with variability around the midpoint further decreasing with age. Ashcraft and Moore interpreted these results as evidence for varying strategies in estimation that changed with age. For example, they reasoned that in Grade 2 students were beginning to use whichever endpoint was closer to the mark to be estimated, whereas use of the midpoint reflected a "midpoint strategy" emerging in Grade 3. Of special interest is the inclusion of an analysis of latency that paralleled the error results. Also, despite their intuition that adults might use quartile points such as 25 or 75 on 0-100, there was no evidence for this pattern in either errors or latency. Ashcraft and Moore ran the same analyses for 0-1000 and similar patterns of error that changed as a function of target position depending on grade were found, but with less differentiation between the grades and only hints of the M-shaped pattern.

To investigate the links between differences in estimation strategy as a function of target position and best-fitting equations, Ashcraft and Moore (2012) also classified students as linear or exponential based on which of the two equations better fit the students' overall response pattern. For Grades 3 to 5, a consistent pattern emerged for all grades: Students with exponential and linear patterns estimated equally well at the origin and endpoints (though Grade 5 linear responders showed less variability overall) but at the midpoint, in all grades, exponential children showed both higher absolute error and more variability. In contrast, linear students in Grades 3 and 4 showed a hint of the M-shaped pattern whereas Grade 5 students clearly showed the M-shaped pattern with low absolute error overall, low variability in error, and less error for the ends and midpoints.

Ashcraft and Moore (2012) interpreted the emergence of an M-shaped pattern of errors as evidence for changes in underlying representation as well as a sign of increased number knowledge, the nature of which is not specified, nor was an empirical test relating specific number knowledge to error patterns reported. They did report, however, that students with estimates best fit by a linear function had higher scores on standardized mathematics tests. Ashcraft and Moore also found evidence for the proportional-judgment model in that some median plots of estimation (Grade 2 on 0-100 and college students on 0-1000) showed the S-shaped curves predicted by proportional-judgment models (e.g., Barth & Paladino, 2011). Accumulating evidence from recent studies investigating children's accuracy trial-by-trial suggests that, in addition to potential changes in underlying representation, differences in *how* students estimate may also be contributing to the observed differences in estimation patterns (Ashcraft & Moore, 2012; White & Szűcs, 2012).

Shortcomings in Current Knowledge

Despite more than a decade of productive research on number-line estimation, two methodological concerns need to be addressed if a comprehensive picture of the development of number-line estimation is to be painted. First, information may be lost or overlooked by averaging across children or individual targets on the line. Second, in most previous studies, children's understanding of number-line estimation has been only narrowly assessed, generally using children's accuracy on the task. Each of these methodological concerns is considered next.

Averaging across individuals and targets. Simon (1975) demonstrated that analyzing how individuals solve problems provides greater insight into the underlying mechanisms of cognition than simply evaluating accuracy. Simon warned that diverse behaviors may be "hidden under a blanket label like 'problem-solution process' even in a very simple task environment," and that to understand problem solving "we must avoid blending together in a statistical stew quite diverse problem-solving behaviors whose real significance is lost in the averaging process" (p. 288). Variants of Simon's approach have been extraordinarily fruitful in leading to a detailed understanding of the development of children's thinking (Farrington-Flint, Vanuxem-Cotterill, & Stiller, 2010; Pressley & Hilden, 2006; Siegler, 2005, 2006).

For example, Siegler (1987) demonstrated the perils of averaging data across trials in children's arithmetic. Initially, students were thought to be solving simple addition problems by counting up from the larger of the numbers being added. When examining both group and individual solution latencies, this account was supported. When Siegler also analyzed students' self-reports, however, a very different picture emerged. Siegler found that students used five different solution procedures, of which counting up from the larger number was only one. Siegler's study illustrated how important insights about solution procedures can be missed when averaging across individuals, as well as across trials. Recent studies aimed at identifying number-line estimation procedures (Ashcraft & Moore, 2012; White & Szűcs, 2012) are a step in the right direction but so far those studies have relied primarily on accuracy. Unless combined with other measures, such as self-reports, accuracy can sometimes be a poor indicator of the procedures used to solve problems (Siegler, 1987; Simon, 1975).

Averaging across individuals and not considering trial-by-trial variability on the number-line estimation task similarly may result in overlooking important information. In the number-line literature, explicit links between representation and procedures are speculative, with only a few exceptions (e.g., Ashcraft & Moore, 2012). The conditions in which students use endpoints and the middle as part of their estimates warrants systematic investigation (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Newman & Berger, 1984; Petitto, 1990; Siegler & Opfer, 2003; White & Szűcs, 2012). Far too little attention has been paid to describing variability trial-by-trial (but see Ashcraft & Moore, 2012; White & Szűcs, 2012) and to individual differences in number-line estimation (Bouwmeester & Verkoeijen, 2012). Rather than assuming that variability is error, characterizing individual variability can shed light on understanding how and when number-line estimation develops (Rouder & Geary, 2014). Thus two important considerations emerge that require further characterization and explanation: (a) variability in individual estimation patterns, and (b) whether and how estimation varies as a function of target (Ashcraft & Moore, 2012; Bouwmeester & Verkoeijen, 2012; Ebersbach, Luwel, & Verschaffel, 2013; Slusser et al., 2013; White & Szűcs, 2012). Bouwmeester and Verkoeijen addressed this problem using data-driven statistical analyses to reveal more variability in estimation patterns than previously found. Their evidence demonstrates that there may be diversity in children's solution procedures in number-line estimation.

Mapping number-line knowledge. In addition to more robust and precise descriptions of individual and task-dependent variability, a clearer picture of the kinds of knowledge that develop in number-line estimation is needed (Ashcraft & Moore, 2012; Bouwmeester &

Verkoeijen, 2012; LeFevre, Lira, Cankaya, Kamawar, & Skwarchuk, 2013; White & Szűcs, 2012). Accuracy is just one source of information that can be used to make inferences about a child's understanding (Bisanz & LeFevre, 1992). Other kinds of evidence, such as students' explanations and justifications of number-line principles, can be used as additional criteria for evaluating what a child knows about number lines. Robustly measuring knowledge using multiple methods has been fruitful in other areas of mathematical cognition (Bisanz, Watchorn, Piatt, & Sherman, 2009; Prather & Alibali, 2009). As important as knowing whether students estimate accurately, is research describing what comprises number-line knowledge, precisely how such knowledge develops, and when it develops. Measuring other indices of number-line estimation knowledge can deepen our understanding of the processes involved.

To help characterize what develops and how, researchers have organized and studied children's mathematical thinking in terms of procedures and concepts. *Procedural knowledge* is "the ability to execute action sequences to solve problems" whereas *conceptual knowledge* is "implicit or explicit understanding of the principles that govern a domain and the interrelations between units of knowledge in a domain" *(*Rittle-Johnson, Siegler, & Alibali, 2001). In other areas of mathematics, procedures and concepts have been shown to develop iteratively and gradually.

As noted, conceptual knowledge can further be categorized into explicit and implicit conceptual knowledge. Explicit conceptual knowledge of the number line may include, for example, knowing about certain characteristics and properties included in the number-line concept, such as knowing that the distance between points on the line must match the arithmetical difference between the corresponding numbers of those points. Implicit conceptual knowledge of the number line is demonstrated for example, in students' accuracy on the task.

Students' can demonstrate the use of number-line concepts when estimating but this may be somewhat separate from the articulation of certain properties or characteristics of the number line as found with explicit conceptual knowledge.

Arguably, it can be difficult to fully disentangle or dissociate procedural and conceptual knowledge but the distinctions among kinds of knowledge can be helpful for understanding students' learning processes and identifying conditions that lead to failure or that support success (Hiebert & Lefevre, 1986). Unfortunately, by measuring only one kind of knowledge, as has been the norm in number-line estimation, important links and relations in the web of number-line knowledge may be overlooked.

Focusing on underlying representation has reduced implicit conceptual knowledge about the number line to one of two kinds: logarithmic and linear representations. This reduction is problematic for several reasons. First, some researchers have found evidence that other equations, such as power or cubic equations, sometimes fit the data better than logarithmic or linear equations (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Bouwmeester &Verkoeijen, 2012; Ebersbach et al., 2008; Rouder & Geary, 2014). Second, focusing almost exclusively on this one measure of implicit conceptual knowledge excludes the wealth of procedural knowledge students seem to draw upon in number-line estimation (Ashcraft & Moore, 2012; Newman & Berger, 1984; Petitto, 1990; Schneider et al., 2008; White & Szűcs, 2012). In addition, processes specific to estimation such as reformulation, translation, and compensation (Reys, 1982, reviewed in LeFevre et al., 1993), as well as rounding and truncation (Lemaire, Lecacheur, & Farioli, 2000), are important kinds of knowledge that may have a place in number-line estimation. For example, students may reformulate the scale of the range on the number-line task. Third, other kinds of explicit conceptual knowledge such as knowledge of proportions may be important, as has been found in other areas of mathematical estimation (LeFevre et al., 1993). Finally, though not explored here, affective components such as tolerance for error, the ability to recognize estimation as useful, and confidence in mathematics may play a role in number-line estimation (LeFevre et al., 1993).

Accuracy, a measure of implicit conceptual number-line knowledge, has been a popular index of number-line knowledge because relations have been observed between individual differences in number-line estimation and other measures of mathematics achievement. For example, researchers have found a positive correlation between an individual's measures of either mean absolute error or linear model fit on 0-100 and 0-1000 lines and overall mathematics achievement (Ashcraft & Moore, 2012; Booth & Siegler, 2006, 2008; Sasanguie et al., 2013; Schneider, Grabner, & Paetsch, 2009; Siegler & Booth, 2004; Träff, 2013). Even when controlling for age, the positive relation holds (Ashcraft & Moore, 2012; Sasanguie et al., 2013; Siegler & Booth, 2004). These striking correlations have led to some speculation about whether number-line estimation may be an index of overall "number sense."

In a longitudinal study with elementary students LeFevre et al. (2013) explored some of the specific kinds of knowledge, such as spatial abilities, numeration or the ability to order symbolic quantities and demonstrate knowledge of place values between 100 and 1000, and calculation, that might be related to, or underlie, number-line estimation. Controlling for grade, sex, and vocabulary, LeFevre et al. found that spatial ability, number-line estimation performance (indexed by the slope of the linear model), numeration, and calculation were all highly correlated with one another.

To examine the longitudinal relations among spatial ability, number-line performance, and calculation, LeFevre et al. used simultaneous path analysis. They found that spatial abilities were positively related to performance on other mathematical tasks including number-line estimation. They also found that spatial ability predicted growth in number-line performance. In contrast to previous work (e.g., Booth & Siegler, 2008), LeFevre et al. did not find that number-line performance measured at the first time point was necessarily more privileged than spatial abilities or numeration in predicting the development of other later mathematics skills. LeFevre et al. (2013) concluded that the number-line task is complex and “growth in performance on the task reflects children’s knowledge of the number system in the specified range in combination with their ability to apply their spatial abilities to create a successful strategy to solve the task” (p. 8). LeFevre et al.’s conclusion speaks to the web of relations between multiple kinds of knowledge, procedural and conceptual, that children may come to rely on in number-line estimation.

As LeFevre et al. (2013) noted, instead of exploring performance on number-line estimation as an index of fundamental “number sense,” “it may be more useful to view the number line task as a measure of children’s ability to skilfully assemble an array of relevant knowledge to perform a complex and (often) novel numerically-relevant task” (p. 9). The opportunity to characterize the ways in which children draw upon their procedural and conceptual knowledge on a task affording inventiveness, such as number-line estimation, is precisely the motivation for looking at the development of flexibility in children’s thinking on the number-line estimation task. Voutsina (2012) noted, by referencing Saxe, Gearhart and Nasir (2001) and Siegler (2001), “the learning and development of meaningful problem solving strategies in mathematics and their adaptive and flexible application depends on a well-connected knowledge and the integration of concepts, facts, and procedures.” (p. 193). Clarifying the kinds of knowledge children draw upon to estimate can shed light on how flexible and adaptive thinking develop in a task suited to such investigations.

The number-line task was used to investigate understanding as a process that occurs within the context of a person and an environment (Baldwin, 1898; Lerner, 2006). Thinking about number-line estimation as part of a developing system encourages reflection about the interactions between multiple levels of that system and about factors both internal and external to the person (Thelen & Smith, 2006; Werner, 1948). Instead of asking whether a child is highly accurate or not in their estimations, I asked "What does the child do?" and "What does the child explicitly know about the number line?" To take advantage of the number-line estimation task as a vehicle for studying flexible thinking more generally, and to describe the graded nature of the knowledge that allows students to exhibit understanding of the task to varying degrees in multiple contexts (Bisanz & LeFevre, 1992; Klahr & Chen, 2011), I articulated the task space associated with number-line estimation (Newell, 1973; Stokes & Baer, 1977).

A Different Approach

The shortcomings in our current knowledge about number-line estimation lead to several questions: How do children solve number-line estimation problems? How do the ways students solve number-line estimation problems change with age or range? What kinds of concepts might students draw upon to estimate? To address these questions I investigated the terrain of number-line knowledge and characterized how this terrain may change as function of grade and range. Because the goal was to provide a "first look," I used a cross-sectional design with students in Grades 2, 4, and 6 to identify and develop measures of age-related psychological phenomena (Schaie & Hofer, 2001) in number-line estimation.

To address shortcomings in the number-line literature about the nature of student's procedural knowledge, I propose a rational task analysis capturing, in detail, the processes a person could use to make a number-line estimate. Next, I propose a framework for characterizing how individual students may think about number-line estimation at three levels of analysis. The most detailed level of analysis is the number-line estimation *processes* students use. How students combine various processes together for a given target reveals their solution *procedures*. The combination of solution procedures students select in relation to targets within a range, I call *tactics*. To capture students' thinking at the level of processes and procedures, I asked students to describe what they were doing as they estimated. Analysis of children's self-reports have been fruitful in leading to a detailed description of how children shift from relying on ineffective to effective strategies at the level of individual problems (Farrington-Flint, Vanuxem-Cotterill, & Stiller, 2010; Pressley & Hilden, 2006; Siegler, 2005, 2006). Finally, I devised a conceptual measure of number-line knowledge to assess the extent to which students understand fundamental properties involved in number-line estimation, such as the idea that there are equal intervals between numbers on the line and how the scale of the line changes.

Task analysis of number-line estimation. I characterized the nature of the number-line estimation task and the processes involved using a rational task analysis. A task analysis, a model of the processes presumably used by problem solvers to complete a task, serves as a starting point for understanding performance (Simon, 1975). The rational task analysis of number-line estimation shown in Figure 3 was informed by the results of previous studies on children's estimation strategies (LeFevre et al., 1993), empirical studies of children's number-line estimation strategies (Newman & Berger, 1984; Petitto, 1990), studies of children's accuracy on individual targets (Ashcraft & Moore, 2012; White & Szűcs, 2012), a review of number-line estimation strategies (Booth & Siegler, 2005; Siegler & Opfer, 2003), eye-tracking studies of number-line estimation (Schneider et al., 2008; Sullivan et al., 2011), and direct observations in an exploratory study of number-line estimation in children with autism spectrum disorders (ASD) (Piatt, Volden, & Bisanz, 2011).

The number-line estimation task is fairly straightforward: One estimates a value on a specified number-line range. A person begins by encoding the range of the line and the numerical value of the target to be estimated. At the end of the task the person makes her estimate, the final output. Both the initial input and final output are assumed to be definite and straightforward, and therefore are not defined further (Newell & Simon, 1972). Initial input, *encode target and endpoints*, and final output, the *final estimate*, are shown in Figure 3.

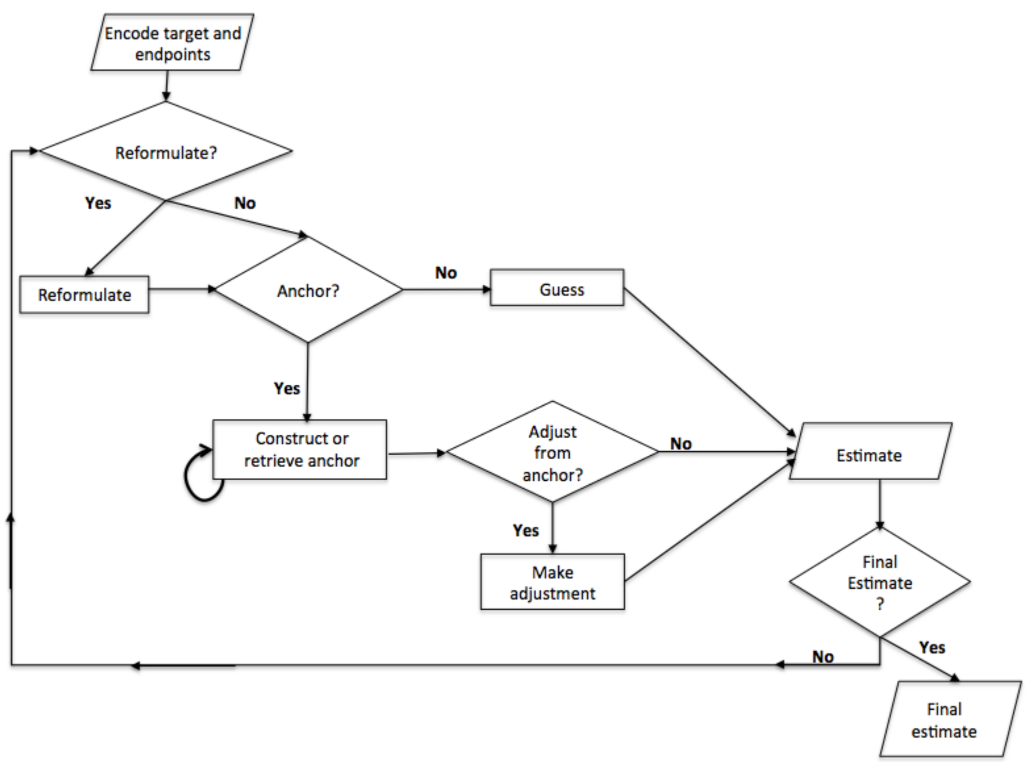


Figure 3. A rational task analysis of number-line estimation. Input and output data are represented by parallelograms, decision points by diamonds, and processes by rectangles.

Next, I define the processes occurring between encoding and final estimation. After encoding, a person may *reformulate* the estimation problem by changing numerical values in one of three ways: rounding (e.g., 265 to 270), truncating or dropping digits (e.g., 265 to 200), or changing the form of the value (e.g., 50 to 5/10) (Reys, 1982, as cited in LeFevre et al., 1993, and in Booth & Siegler, 2005). In the case of number-line estimation two numerical values might be reformulated: the target to be estimated and the range or scale of the problem. In Figure 3 only the general process of reformulating is shown.

After reformulating or not, children either *anchor* to a point, or several points, that serve to constrain subsequent operations, or they simply *guess*. If a child anchors, three processes can be invoked: (a) iteratively anchoring to establish further points of constraint, (b) making a final *adjustment* from one or more anchors to establish an estimate, and (c) deciding that the anchor is "close enough" to serve as an estimate. Early researchers of number-line estimation outlined both anchoring and adjustment processes (Newman & Berger, 1984; Petitto, 1990). Recent studies have confirmed both processes by assessing children's eye movements (Schneider et al., 2008) and gestures (Segal, 2011) during the task. In addition, patterns of latency and error are also consistent with the use of anchors and adjustment (Ashcraft & Moore, 2012; White & Szűcs, 2012). The rational task analysis shown here, however, is the first to represent the organization of previously reported estimation processes in a comprehensive model of number-line estimation.

To illustrate how a student's estimates might be understood through the lens of the task analysis, consider the following case: A student sees the number line from 0 to 100 and the target to be estimated is 67. The student may move through the processes of the task analysis in a number of ways. First, she may see (and encode) the target and the end points and not change the numerical values, thereby not invoking a reformulation process. She may then anchor to a point near 67, such as 50, the middle or halfway point on 0-100, and then adjust from 50 by counting by fives to 65. Thinking that this value is close enough to the target number, she may not invoke any further processes and indicate her final estimate.

A student might estimate the location of 67 on a 0-100 line in other ways. An alternative path through the task might be to encode the target and the end points but then decide to "look for 70", thus reformulating the problem by rounding the target from 67 to 70. Next, this particular student may anchor to the 100-end of the number line, adjust from 100 by counting down by tens to arrive at 70 and stop, thereby establishing his final estimate. Another student may take a similar tack but reformulate both the target and range of the line. That is, in estimating the location of 67 on a 0 to 100 number line, after encoding the problem, the student might "look for 7 out of 10," essentially rounding 67 to 70 and transforming the target from 70 to 7 and the range from 100 to 10. The student might then count from 0 by ones to 7. The task analysis demonstrates that there are many ways to make an estimate.

The task analysis is organized to capture broad families of processes, such as anchoring and adjusting, and also the many ways in which students may link processes together, leaving room for further specification within a process. For example, one student could anchor first to 0 and adjust by counting up by ones to 67. The model is broad enough, however, to capture a similar set of processes but with different values where a different student could anchor first to 100 and adjust by counting down by tens to 60 and then adjust by counting up by ones to 7 to arrive at 67. Another example: A student might retrieve anchors for 50 and 75 (half and three-quarters) and then adjust "between those two" to get "about 67". These examples illustrate that the task analysis should capture generally and accurately the range of moves students may make when estimating.

Characterizing problem-solving operations. Given a rational task analysis to guide analysis of performance on each estimation problem, the next step was to develop a way to characterize problem solutions within and across problems. The Levels of Problem-Solving Operations taxonomy is a framework for operationalizing levels of problem solutions across the number-line task specifically (Table 1), but it may be generalized to other kinds of problem-solving tasks. In the context of number-line estimation, *process* refers to a component solution process such as anchoring to an endpoint or adjusting by counting. Solution *procedure* refers to the particular solution process or combination of component solution processes a child uses when estimating a target, such as anchoring to the midpoint and then counting up by ones to locate 67. Procedures shed light on students' thinking on a trial-by-trial basis. In number-line estimation analyzing only individual trials fails to account for how individual students select procedures across a number-line range. Understanding how individuals select procedures as a function of target is important because certain targets should afford, or evoke, possible procedures over others depending on what the student knows (Gibson, 1979/1986), as has been shown in previous work (Ashcraft & Moore, 2012; Newman & Berger, 1984; Petitto, 1990). *Tactics* capture the patterns of solution procedures students select across targets within a range (Table 1). By combining the details of the task analysis with the taxonomy of Levels of Problem-Solving Operation, I am able to characterize the procedural knowledge associated with number-line estimation on both a trial-by-trial basis, as well as for an individual within a range, and then investigate how students adjusted their tactics across ranges.

Table 1

Levels of Problem-Solving Operations

Level	Operation	Definition	Scope of Analysis in Number Lines	Examples
I	Process	A component; part of a solution procedure on a single trial	Within a trial for a particular target	For a particular target, anchoring to 0 or adjusting by counting up
II	Procedure	The individual process or combination of processes used to solve a problem on a single trial	Within a trial for a particular target	Locating 67 on 0-100 range by finding midpoint (first process) and then adjusting the answer (second process) linked together as a procedure
III	Tactic	The combination of solution procedures used to solve problems across several trials within a task	Across trials within a specified number line	Locating most targets within a 0-1000 by using only the endpoints as anchors to estimate (an ends-only tactic)

Processes. Understanding the processes used in number-line estimation, or any task, creates a foundation for understanding how children's thinking about the task develops over time (Siegler, 2005; Simon, 1975). Only a few researchers have investigated *how* students estimate on number-line tasks by assessing processes or procedures either directly (Newman & Berger, 1984; Petitto, 1990) or indirectly (Ashcraft & Moore, 2012; Schneider et al., 2008, Sullivan et al., 2011; White & Szűcs, 2012). As introduced in the levels of problem-solving operations, processes form the most basic unit of problem solving. The task analysis of number-line estimation (Figure 3) illustrates number-line estimation processes at a general level. Within each general process exist specific types of that process. For example, there are many specific types of anchors a child could use to estimate, such as anchoring to the ends of the number line or to the midpoint. For each process in the task analysis, a comprehensive list was generated of the specific types of each process (Appendix A) and later used to code the data. Processes formed the foundation for thinking about procedures, the next level of problem-solving operations.

Procedures. The combination of processes students used to estimate a target form a procedure. For example, to locate 67 on 0-100, a student might first invoke the process of anchoring to 50, then use an adjusting process and move upward "about 17 spaces." The combination of this anchor and adjustment forms a procedure. Based on the literature that informed the task analysis as well as my own observations of students estimating on a number line, a few kinds of procedures were suggested. Just as Newman and Berger (1984) organized rules according to a level of sophistication, I propose that procedures may be ordered to capture whether students are using procedures that indicate anchoring to only part or to the whole line. The least advanced procedure may be to *guess*, neither anchoring nor adjusting, and therein not explicitly indicating any sense of orienting to the line. A more advanced procedure, in contrast, may be a procedure by which students anchor to a landmark such as the 0-end, middle, or high end; procedures that could be called *ends* (0-end or high end), and *middle*. Finally, an even more advanced procedure may involve anchoring to *proportions* because to use proportions the student would implicitly think of the line as a whole and make a proportional judgment (as suggested by Barth & Paladino, 2011). How students then assemble their procedures across targets within a range would reveal their tactic for that range.

Tactics. The collection of procedures a student uses across targets for a particular number-line range is a tactic (see Louis Lee & Johnson-Laird, 2012, for use of the term when adults solve a series of related problems). Considering tactics in number-line estimation is particularly important because of the different patterns of procedures that different targets may evoke (Ashcraft & Moore, 2012; Newman & Berger, 1984; Petitto, 1990; Schneider et al., 2008; White & Szűcs, 2013). I propose a ladder of tactics, grounded in previous literature and ordered to suggest a developmental sequence through which students may move as their knowledge about number lines progresses from immature to advanced levels (Figure 4). The tactical ladder is organized to suggest that development of number-line estimation reflects increasing differentiation and hierarchical integration of number-line knowledge including knowledge about landmarks, scale, proportions, and equal intervals. The tactical ladder specifies the kinds of procedures children use as a function of target positions (Ashcraft & Moore, 2012; Schneider et al., 2008; White & Szűcs, 2012). Here, as previous researchers have done, the range is divided into five segments (for their exact specifications see Ashcraft & Moore, 2012; Schneider et al., 2008). I divided the line into quintiles. For example, for 0-100, the first quintile, called Origin, is from 0-20. The second quintile, from 21-40 includes the one-quarter mark and is called First Quarter. The third quintile, from 41-60 includes the middle, and is termed Middle. The fourth quintile, from 61-80, includes the three-quarter mark and is called Third Quarter. The final quintile is from 81-100 and is called End.

Development

Skill Level	Tactic	Quintile on Number Line				
		Origin	First Quarter	Middle	Third Quarter	End
Expert	Pure Proportions	*Proportions*				
Advanced	Proportional Landmarks	*Ends*	*Proportions*	*Middle*	*Proportions*	*Ends*
Basic	Landmarks Middle	*Ends*	*Middle*			*Ends*
Emerging	Ends only	*Ends*				
Immature	Guessing	*Guess*				

Figure 4. Proposed number-line estimation tactics.

At the bottom of the proposed tactical ladder is guessing, hypothesized to be associated with minimal knowledge about the number line. If relying on a *Guessing* tactic, students guess across most quintiles of the number line. Students climb to the next rung on the tactical ladder when they recognize the number line as a bounded task, to some extent, and are able to orient to and use the ends of the line. If relying on an *Ends-only* tactic, students predominantly rely on one or both ends to estimate across quintiles. Evidence for use of ends as the earliest kind of anchor and foundation for one of the less mature tactics comes from the few studies of children's thinking in number-line estimation (Ashcraft & Moore, 2012; Bouwmeester &Verkoeijen, 2012; Newman & Berger, 1984; Petitto, 1990; Rouder & Geary, 2014; Schneider et al., 2008) and speculations about possible strategy-use in number-line estimation (Ashcraft & Moore, 2012;

Bouwmeester &Verkoeijen, 2012; Barth & Paladino, 2011; Siegler & Opfer, 2003; Siegler & Booth, 2005; Slusser et al., 2013; White & Szűcs, 2012).

As students reach the next rung, they appropriately use the two ends of the line to estimate at the origin and end and have learned to orient to the middle at least for targets in or near the middle quintile. This tactic, *Landmark Middle*, is the most frequently alluded to "tactic" in the literature (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Newman & Berger, 1984; Siegler & Opfer, 2003; Petitto, 1990; Siegler & Booth, 2005; Slusser et al., 2013). Evidence for use of the middle comes from studies of children's self-reports on number-line estimation (Newman & Berger, 1984; Petitto, 1990), eye-tracking studies showing that children orient to the middle more with age (Schneider et al., 2008) and that adults orient to the ends and middle (Sullivan et al., 2011), and use of particular anchoring strategies in number-line estimation (Ashcraft & Moore, 2012; Bouwmeester &Verkoeijen, 2012; Barth & Paladino, 2011; Siegler & Opfer, 2003; Siegler & Booth, 2005; Slusser et al., 2013; White & Szűcs, 2012). Students using the *Landmark Middle* tactic are hypothesized to have a basic level of number-line knowledge in that they recognize that the number line can be thought of as a bounded task with the halfway point being a useful anchor.

Use of distinct landmarks in each of the five quintiles is the next rung on the ladder. A *Proportional Landmark* tactic is hypothesized to be students' continued use of ends and midpoint, but also the use of proportions around the first and third quarters of the line. Finally, the top-most rung or tactic, *Pure Proportions*, reflects the idea that the number line can be understood and represented as perfectly linear with each number and spatial position represented as a proportion (Siegler & Opfer, 2003). Students using a Pure Proportion tactic consistently operate with Heeffer's (2011) definition of the number line as a representation in which points on the line match the arithmetical difference between corresponding target numbers.

Though the tactical ladder is shown as a series of discrete levels, two caveats must be made. First, these levels are neither stage-like nor discrete. Tactics are shown as distinct from one another for clarity but there may be several strata or sub-levels within each tactical level. Moreover, levels may be permeable. For example, in moving from *Landmark Middle* to *Proportional Landmarks*, students might use the lower half of the line first, as a kind of in-between-level that is no longer just landmark middle but is not the fully developed proportional segmenting level.

Adjusting tactics. Finally, researchers have shown that children choose solution procedures in an adaptive way, not only in the realm of computational estimation (Lemaire et al., 2000) but on other kinds of mathematical tasks (Siegler & Jenkins, 1989; Lemaire & Lecacheur, 2002) and in other domains such as reading (Siegler, 2005). For example, Lemaire and colleagues used accuracy and speed as indices of strategy effectiveness in computational estimation and found that students chose solution procedures that increased accuracy and reduced problem-solving time thereby exhibiting adaptive strategy selection. Extending the idea of adaptivity to number-line estimation, I explored how students chose solution procedures with respect to target quintile for two number-line ranges, to construct tactics for each line, 0-100 and 0-1,000. Then, once students' tactics were classified on the two ranges, I explored whether and how students adjusted their tactic from the 0-100 to the 0-1000 number line relative to how much change in error they accrued when moving from 0-100 to 0-1000. Children were expected to adjust their solution operations adaptively with a change in number-line range so that they either maintained or improved accuracy. Moreover, because adjusting tactics is an extension of the idea that children selectively choose solution procedures to be adaptive (Lemaire et al., 2000; Siegler & Jenkins, 1989), older children were expected to adjust their tactics more effectively than younger children. Notably, some children may fail to adapt.

Self-reports. Students' self-reports were used to capture the level of detail needed to learn about students' processes, procedures, and tactics on number-line estimation. Self-reports have been shown to be valuable in characterizing children's thinking (Siegler, 1987; Simon, 1975). Students' descriptions of their thinking while estimating may cause children to change their thinking over time, an effect called reactivity (Fox, Ericsson, & Best, 2011; Russo, Johnson, & Stephens, 1989). Russo et al. (1989) noted that verbal protocols have generally been shown to slow processing but not fundamentally to change thinking. Nevertheless, whether thinking aloud changes children's thinking in number-line estimation is an empirical question. In a separate study of number-line estimation with Grade 6 students, thinking aloud was not found to affect children's accuracy on number-line estimation (Appendix B).

Establishing the veridicality of children's self-reports is also important. In a concurrent think-aloud paradigm such as the one used here, it is difficult to assess veridicality without additional data on children's processes using, for example, eye-tracking (Russo, Johnson, & Stephens, 1989). All sessions were video recorded so that both verbal and gestural information, potentially an important kind of process data, could be assessed. In addition, as recommended by Russo et al. (1989), a theory-based approach was used along with a detailed task analysis to serve as external criteria in the classification protocol of self-reports. To classify children's self-reports, both verbal and behavioural evidence were used, but in cases of mismatch, which happened rarely, the coder relied primarily on her observations of the child's behavior.

Explicit conceptual number-line knowledge. I designed a number-line knowledge task to measure students' explicit conceptual number-line knowledge for two reasons. First, using multiple methods to assess understanding is a more robust and informative approach than measuring understanding with one method alone, such as accuracy (Bisanz et al., 2009; Bisanz & LeFevre, 1992). Second, information from the explanation task was used to further explore and understand children's reported processes and procedures, and to help shed light on why children may or may not have successfully adapted their estimation solution operations within a range and across tactics. Students completed the number-line knowledge task at the end of the final number-line session.

The number-line knowledge task was designed to measure children's explicit conceptual knowledge about number lines (K_{num}) including knowledge about numerical categories (Laski & Siegler, 2007), important aspects of the definition of the number line such as scale invariance and equal intervals (Heeffer, 2011), and important landmarks including the midpoint (Ashcraft & Moore, 2012; Bouwmeester & Verkoeijen, 2012; Siegler & Opfer, 2003) and proportions like one-quarter and three-quarters (Barth & Paladino, 2011; Slusser et al., 2013).

Use of an explicit conceptual measure of the number allowed for exploration into (a) the kinds of implicit conceptual knowledge that are often taken as reflections of implicit conceptual knowledge, (b) whether there is variability in the kinds or degree of explicit conceptual knowledge students have despite demonstrating similar estimation patterns within a range, and (c) the relations among math achievement and number-line estimation as indexed by both implicit and explicit measures of conceptual knowledge.

A 14-item questionnaire was designed to include some additional estimates ("Where would you put 50 and 60 on the 0-100 line?") as well as probes about number-line concepts such as basic knowledge ("Where are the small, medium, and large numbers on the line?"), scale changes ("What number goes in the middle if this endpoint changes from 100 to 1000?") and equal intervals ("Does the space between 90 and 91 have to be the same as the space between 10 and 11, or can it be bigger or smaller?"). Students answered several questions about the general structure of the 0-100 number line followed by several questions about anchors on the 0-100 number line. Next came questions designed to determine whether the student understood the nature of equal intervals on the number line and a few questions about what happens to certain numbers when the scale changes from 0-100 to 0-1000. Finally, all students were asked about three proportions on the number line: quarter, half, and three-quarters. Details of the task and coding scheme are in Appendix C. Student's knowledge about number lines was indexed by summing across nine items to reveal their K_{num} score. K_{num} scores could range from zero (0s on all nine items) to nine (1s on all items).

Summary and Preview

Two shortcomings in current knowledge about number-line estimation need to be addressed. The first is that variability in individual estimation patterns and the variability related to task affordances must be described and explained. The second shortcoming is that number-line knowledge has been only narrowly assessed, generally based solely on students' accuracy on the task. To tackle these issues, a multi-method approach to measurement was used and an emphasis was placed on identifying variability at the level of targets and individuals. Children provided self-reports while estimating targets on 0-100 and 0-1000 number lines. Students estimated targets on the number-to-position task, used by previous researchers (e.g., Barth & Paladino, 2011; Booth & Siegler, 2005; Siegler & Opfer, 2003). To ensure the task was clear to students, task instructions were slightly modified to include the explanation of a number line being like a ruler (based on Petitto, 1990). Of note is that no orientation to, or feedback about, the middle of the line was given.

In this study, both implicit and explicit conceptual knowledge were measured, as was procedural knowledge. As in most previous studies, a measure of implicit conceptual number-line knowledge, students' accuracy, was assessed. Results from the number-to-position task were analyzed to ensure comparability to previous studies and to verify that self-reports did not drastically change estimation patterns. Explicit conceptual knowledge was measured with a novel measure, the number-line knowledge task, designed to identify and probe students' understanding of key features and principles of the number line.

Finally, procedural knowledge was measured by coding self-report data from students' estimates in terms of processes, procedures, and tactics. A task analysis of number-line estimation was proposed to specify and code the solution processes children might use as the basis for evaluating students' procedures on the task. Classifying students' procedures allowed for the examination of variability as a function of target. Tactics emerged from classifying patterns of procedures and allowed for an examination of individual variability. The purpose of measuring procedural knowledge was to address the following questions: What procedures do children use to solve number-line estimation problems? How does the selection of procedures change with increasing age and across two ranges? What tactics do individual children use across each of the two ranges? How does the use of tactics change with age and across the two ranges? How do individual differences in number-line estimation relate to students' accuracy within a range?

With data from multiple measurements of number-line knowledge, the following questions were addressed: What is the relation between implicit and explicit conceptual knowledge on the task? What is the relation between individual children's tactics, a measure of their procedural knowledge, and their level of explicit number-line knowledge? Finally, are more advanced procedural and explicit conceptual number-line knowledge, indexed by tactics and the explicit number-line knowledge task respectively, positively related to scores on standardized tests of mathematics achievement?

To answer these questions and capture potential changes in number-line estimation abilities and degrees of knowledge, students in Grades 2, 4, and 6 participated in four testing sessions. To compare results from this study with previous studies that used standardized assessments, students completed a battery of standardized tests in the first two sessions. In the third session students estimated 20 targets on two typical ranges, 0-100 and 0-1000, and provided self-reports on half the trials. In the fourth session students estimated five targets on each of four atypical (Booth & Newton, 2012), or non-canonical, ranges (such as 0-80 and 0-531) while reporting their thinking and then completed the explicit number-line knowledge task. These data were used to explore the levels of problem-solving operations in number-line estimation. A framework was proposed for thinking about problem-solving operations in number-line estimation. That framework allowed for (a) characterization of how individual children think about the task, (b) description of how thinking changes as a function of grade and range, and (c) exploration of the relations among procedural and explicit and implicit conceptual procedural knowledge. By better characterizing the kinds of knowledge that develops in number-line estimation, a foundation is laid for exploring the best ways to support the teaching and learning of number-line estimation. Finally, the method used here illustrates a new way of characterizing flexible thinking more generally by characterizing and identifying patterns of individual variability as a task changes both trial-by-trial, and from one number-line range to another.

Method

Elementary school students participated in four sessions, each lasting approximately 30 minutes. In Sessions 1 and 2 students completed standardized tests from the Woodcock-Johnson III (Woodcock, McGrew & Mather, 2001) to characterize their general cognitive abilities. Between Sessions 1 and 2 the median number of days was 10. In Session 3, students estimated targets on typical ranges. In Session 4 students estimated targets on atypical ranges and completed the explicit number-line knowledge task. Between Sessions 3 and 4 the median number of days was seven. The results from atypical ranges are beyond the scope of this study but are described because estimating on those ranges may have affected performance on the explicit measure of number-line knowledge.

Participants

In each of three grades—2, 4, and 6—24 typically developing children (12 girls) were recruited from public elementary schools in small communities surrounding a Canadian city. With permission from the school district and school principals, children were recruited through classroom teachers. All test sessions took place in quiet rooms in the schools. Grade 2 students ranged in age from 7;05 to 9;00 (years;months) with a mean of 7;11, Grade 4 students ranged from 9;4 to 10;4 with a mean of 9;10, and Grade 6 students from 11;2 to 12;7 with a mean of 11;9. Data about socioeconomic status and ethnicity were not collected but children were drawn from three communities (A, B, and C) in which the majority of adults had at least a high school diploma (83%, 73%, 57%, respectively), and fewer than twenty percent had a bachelor's degree or a higher degree (14%, 8%, 3%). In all three communities, most residents self-identified as Caucasian (96%, 97%, 96%). Median household income differed ($98K, $67K, $42K; Statistics

Canada, 2006). Grade level was confounded with community: All data for Grades 4 and 6 came from Community A, and data for Grade 2 came from Communities B and C.

The Human Research Ethics Board at the University of Alberta approved the research. The author and a fourth-year undergraduate psychology student served as experimenters. Data from an additional four children were collected but not included in the final sample due to experimenter error.

Standardized Tasks and Measures

In Sessions 1 and 2, students completed standardized subtests from the Woodcock-Johnson III (Woodcock et al., 2001). Half the children completed the tests of Brief Intellectual Ability (BIA) in Session 1 and then completed the Understanding Directions, Math Fluency, and Applied Problem Solving subtests in Session 2. The order was reversed for the remaining children. The Woodcock-Johnson III (WJ-III) is a battery for measuring the cognitive abilities and school achievement of school-aged children and young adults. The WJ-III includes Canadian content and norms. Updated norming procedures were used in the most recent estimate of reliability indices for the U.S. sample (Woodcock, McGrew, Schrank, & Mather, 2007). Because no significant differences were found between U.S. and Canadian samples (Ford, Swart, Negreiros, Lacroix, & McGrew, 2010), reliabilities for U.S. samples (McGrew, Schrank, & Woodcock, 2007) are reported below. Only U.S. median reliabilities for ages 7 to 12 years for the subtests used in this study are available.

Brief Intellectual Ability (BIA). The test of Brief Intellectual Ability (BIA) is composed of three subtests: Verbal Comprehension, Concept Formation, and Visual Matching. This test is used for quickly and accurately assessing general cognitive ability. Verbal Comprehension is an untimed test requiring students to label pictures of objects, produce antonyms and synonyms, and complete verbal analogies. Concept Formation, also untimed, is a visually presented test of rule-based categorization, using shapes, in which the student is required to articulate the correct rule for categorizing a shape on each trial. Visual Matching is a timed test in which the child circles pairs of matching numbers within a defined set of numbers. Reliability, indexed by Cronbach's alpha, is .90 for Verbal Comprehension, .94 for Concept Formation, and .97 for the BIA, overall. Reliability for Visual Matching was calculated using a Rasch analysis procedure and is .88.

Understanding Directions. Scores on this subtest were used to index language comprehension for each child. The test requires the student to listen to a sequence of instructions and then follow the directions with a motoric response (pointing). Split-half reliability for this test is .81.

Math Fluency. The math subtests of the WJ-III are often used as indices of math abilities (e.g., De Smedt, Holloway, & Ansari, 2011). The Math Fluency subtest requires the child to answer as many single-digit addition, subtraction, and multiplication problems as possible within three minutes. Reliability, calculated using a Rasch analysis procedure, is .95.

Applied Problem Solving. In this untimed subtest students must verbally answer increasingly difficult calculation problems presented orally or visually. Split-half reliability for this test is .92.

Number-line Estimation Task

The number-line estimation task consisted of six number-line ranges over two sessions: two *typical* ranges adapted from the number-to-position estimation procedures of Booth and Siegler (2006) and Barth and Paladino (2011) in Session 3, and four *atypical* ranges in Session 4. Number lines were presented with a specially designed computer application (app) on a touch screen (iPad). Sessions were video recorded. Accuracy, solution-process behaviour, and self-report data were collected.

Number-line stimuli. Two typical number-line ranges were used in Session 3: 0-100 (Barth & Paladino, 2011) and 0-1000 (Booth & Siegler, 2006) (Table 2). For 0-100, two targets were selected from every decade; for 0-1000, two targets were selected from every century. Both ranges were presented as two sets of 10 targets each with a short break between sets. Children were asked to give self-reports on each problem in the second set. In Session 4, students estimated and gave self-reports on five targets in each of four atypical number-line ranges (Booth & Newton, 2012) (Table 2). Two presentation orders of targets within each range were counterbalanced across students with grade and gender.

Table 2

Stimuli for Number-line Estimation Tasks

Session	Type of Number Line	Range	Within-session Order	Self-report	Order of Targets [a]
3	Typical	0-100	Set 1	no	21, 48, 61, 94, 57, 36, 88, 17, 8, 76
			Set 2	yes	96, 33, 4, 52, 81, 42, 14, 67, 25, 72
		0-1000	Set 3	no	3, 907, 545, 721, 391, 184, 613, 409, 835, 215
			Set 4	yes	475, 19, 158, 654, 240, 760, 502, 992, 325, 805
4	Atypical	0 -80	Set 1	yes	62, 5, 22, 75, 42
		0 -74	Set 2	yes	4, 70, 54, 17, 36
		0 -220	Set 3	yes	114, 13, 51, 207, 169
		0 -531	Set 4	yes	255, 32, 143, 499, 397

[a] Targets were ordered unsystematically with the constraint that no targets from the same decade appeared in adjacent positions and that trials did not appear in ascending or descending order. Target order was reversed within a set to create two presentation orders.

Presentation. The number-line estimation task was presented as a computer iPad application (app) developed in collaboration with a Computing Science graduate student at the University of Alberta. Number-line stimuli were presented as a straight line on the touch screen, 18 cm in length, with 0 marked at the left end and the endpoint of the number-line range (e.g., 100 or 531) marked at the right end. The target number was centered 2.5 cm above the line. To make their estimates on the iPad, students used a Kuel H10 highly sensitive stylus.

Most research on number-line estimation has been conducted with pencil and paper or a desktop computer (Ashcraft & Moore, 2012; LeFevre et al., 2013; Newman & Berger, 1984 Schneider et al., 2008; White & Szűcs, 2012). In a separate study we demonstrated that students performed comparably on number-line estimation on the iPad and with pencil and paper (Piatt, Coret, Choi, Volden, & Bisanz, 2014; see Appendix D for details).

Instructions for administration. Experimental procedures were adapted from Barth and Paladino (2011), Booth and Siegler (2006), and Pettito (1990).

Typical number-line ranges*.* The experimenter explained that the purpose of the study was to explore how students think about number lines and then obtained their assent. Students were oriented to the video camera, iPad, and stylus. Next, students were introduced to the number-line task and asked to show the experimenter, by tapping with the stylus, where they thought the number at the top of the screen would go on the number line. The experimenter explained, while showing the child a ruler and gesturing to the ends of the ruler and numbers in between, “A number line is like a ruler. It has numbers and marks at both ends and all along it.” The experimenter then started the iPad, showed the child a screen with a 0-100 line, and said, “This is a number line too, but on this number line only the numbers at the ends are written in.” A new screen appeared with the numbers on each end enlarged 200% and then scaled down to the normal size. This screen, meant to draw attention to the current range, appeared at the start of every new number-line range. During this screen the experimenter said:

> The flashing red numbers remind us what number line we are working on. Now we are working on a number line from 0 to 100 (or *R*, on later sets). Remember, tap quickly, like this *(experimenter demonstrates)*, and make sure your hand is not touching the screen. Let’s say I put the answer for the number at the top of the screen (100 shown on screen) here *(experimenter marks near but not on 100)*, but that is not where I think the answer is. It’s okay to make another mark like this *(experimenter marks on 100)*. When you tap

> your best answer, tell me "done", and then I will tell you it's okay to press the OK button to lock in your answer *(experimenter demonstrates pressing OK button above the mark)*.

Finally, students practiced using the stylus, making marks and pressing the OK button by showing the experimenter where the practice targets, 0 and 100, were on the 0-100 number line. Just prior to starting the set, the experimenter reminded the student: "Remember, your job is to show me, by tapping with the pen, where you think the number at the top of the screen goes on the number line. Make your *best* answer as quickly as you can and then tell me 'done' when you have your answer and I will tell you to press 'ok' and lock in your answer." Students then estimated 10 targets on a 0-100 number line (Table 3).

After completing the first set, a screen appeared with a video of some cheerful music and animated dancing rabbits holding a sign that said "Nice Job!" The experimenter told the student:

> Good Job! Now we are going to do something a little different. This time I am going to ask you to think out loud and tell me in a loud and clear voice what you are doing in your head as you decide where to put numbers on the number line.

To practice thinking out loud, the student was asked to draw a house on a commercially available app (Drawing Pad) on the iPad, and told "I want you to draw a house and at the same time as you are drawing, tell me in a loud and clear voice, what you are doing as you draw." If the student seemed confused or said that he or she did not understand, the experimenter demonstrated by drawing the base of a house and two walls and saying. "I am drawing the walls of the house," and then she drew the roof saying, "and now I am drawing the roof. " The experimenter erased her work and handed the stylus to the student, "Now you try."

If the student started drawing but did not report on his or her thinking, the experimenter prompted, "Tell me what you are thinking as you draw." All children were able to clearly report their thinking as they drew a house.

Just prior to beginning the second set of 10 targets on a 0-100 number line, the experimenter reminded the student:

> Remember, tell me in a loud and clear voice, just like we practiced with the house, what you are thinking about as you decide where to put the numbers on the number line. Tap your best answer as quickly as you can and tell me done. Once you say "done," you cannot tap or touch the screen. I will ask you a question after each answer. Do your best to tell me what you can, then I will tell you to press the OK button.

At the end of the trial, the experimenter always asked the student to describe her thinking, to obtain a self-report of solution processes (e.g., Siegler & Stern, 1998). If the student did not explain her thinking while estimating, or the explanation was unclear, the experimenter said, "How did you decide to put your answer there?" On the next trial the experimenter reminded the student to describe her thinking while estimating. If the student clearly described his thinking while estimating but the answer was unclear, the experimenter asked, "Did you do anything else?" In contrast to previous studies of number-line estimation in which self-reports were recorded for just a few trials (Newman & Berger, 1984; Petitto, 1990), self-reports were collected on 10 targets from across the entire range for both 0-100 and 0-1000. Targets used with the verbal self-report protocol were the same for all children.

The same instructions were given for the 0-1000 number-line ranges. Because a new range was introduced students were oriented to a 0-1000 number line with the flashing red numbers and were asked to show the experimenter, on a line with both endpoints visible, where 0 was on the line and where 1000 was on the line. As with 0-100, children made their estimates for the first ten targets silently and were reminded to think aloud for the last ten targets. Students were given general praise and encouragement throughout the session but were not given feedback on individual trials. At the end of the session students were thanked and reminded they would be seen one more time to play the number line game on the iPad.

Atypical number-line ranges. The fourth session began with reaffirming the student's assent and re-orienting the student to the video camera and iPad. For all trials on the atypical ranges students were instructed to think aloud. Before starting with the number-line app, to practice thinking out loud, students drew a car and at the same time, explained their thinking.

At the start of each new, atypical number-line range, the same orienting procedure described above was used. Children estimated five targets (Table 1) on each of the unconventional ranges. Each range was followed by a brief break. Ranges were always presented in the same order.

Data collection. On all trials, accuracy was recorded using the iPad app. Any additional marks, such as drawing a succession of lines to count units on the screen, were also recorded by the app. Visible strategy use was recorded on every trial (classification details are outlined in the analysis section). Forty verbal protocols were collected for each student: 10 from targets across each of the two typical ranges and five from targets across each of the four atypical number-line ranges. All trials were videotaped for later review of behaviour and self-report as necessary.

Explicit Number-line Knowledge Task

Following the last atypical number-line range, students were given the explicit number-line knowledge task. First students were asked to explain a 0-100 number line to another student: "Sam is a student at another school learning about number lines. How would you explain to Sam what the number line is and how it works?” Second, students were asked a series of 14 questions designed to probe their understanding of explicit number-line concepts. Items were designed to assess the students’ understanding of essential concepts and principles of the number line such as scale (What happens to 50 and 60 when this end changes from 100 to 1,000?), equal intervals (Is the space between 10 and 11 the same as between 90 and 91?), and proportions (Where is ¼ on the line?) (Table C1). One question was deleted because all students answered it correctly. Some items were combined (see Appendix C for details) resulting in a total of nine items that captured children’s explicit conceptual number-line knowledge.

This task was not piloted and as a result was slightly modified as the study progressed. The first version, given to the first five students, included nine questions designed to reveal children’s understanding of number lines and eight estimates of potential landmark targets (50, 30, 60, one-quarter, half-way, three-quarters on a 0-100 number line). In this version questions about equal intervals were not clear enough to reveal whether children understood what happens to the intervals between numbers as the scale changes from 0-100 to 0-1000. To clarify children’s understanding of how intervals change as the range changes, another question and two more landmark estimates were included (50 and 60 on a 0 to 1000 range). The remaining students (n = 67) were given this second version.

The explanation task always took place at the end of the final session because children’s reflections and explanations of the number line could have resulted in changes to their understanding (Siegler, 2002). At the end of this final session, students were thanked for their help and given the opportunity to ask questions about the study.

Results and Discussion

Analyses of the data are presented in six sections. First, because students in the sample came from communities with different SES levels, achievement scores were tested to assess whether achievement levels were confounded by differences in SES across the communities. Second, because this study used a specially designed computer app and self-reports, accuracy data from this study were compared to accuracy in past studies. Third, relations between the newly devised measure of explicit conceptual number-line knowledge and implicit conceptual number-line knowledge were explored. Fourth, a method for identifying students' processes, procedures, and tactics for number-line estimation was tested and differences in procedures and tactics were analyzed as a function of age and range. Fifth, the extent to which the procedural and conceptual measures were related with measures of math achievement was examined. Finally, to take advantage of the number-line task as a window into students' adaptive thinking, analyses were conducted to explore whether and how students adjusted their number-line tactics across ranges.

Summary of Achievement Scores

The mean standard scores on the WJ-III for the students in this study in Grades 2, 4, and 6 are shown in Table 3. Recall that grade was confounded with community: Children in Grade 2 came from communities with lower median incomes and education compared with students in Grades 4 and 6 (Statistics Canada, 2006). A one-way analysis of variance was conducted for each subtest. Grade-related differences in standardized scores were negligible for Brief Intellectual Ability and for Understanding Directions. Performance on Math Fluency varied somewhat with grade, $F(2,69) = 2.99$, $p = .06$, $\eta_p^2 = .08$, primarily because the mean for Grade 4 was lower than for the other two grades, *ps* = .04. Performance on Applied Problems also varied with grade, $F(2,69) = 8.17$, $p = .001$, $\eta_p^2 = .19$, with Grade 6 scores intermediate (ps = .05). Thus the mathematics ability of Grade 2 students was not likely to have been affected adversely by demographic differences in household income or parental education.

Table 3

Mean Standard Scores for Each WJ-III Subtests by Grade

Grade	Brief Intellectual Ability *M* (*SD*)	Understanding Directions *M* (*SD*)	Math Fluency *M* (*SD*)	Applied Problems *M* (*SD*)
2	105 (11)	106 (9)	99 (11)	109 (13)
4	101 (11)	103 (16)	91 (15)	94 (16)
6	103 (11)	101 (8)	99 (11)	101 (8)

Note. Age-based standard scores calculated with $M = 100$ and $SD = 15$ using WJ-III Normative Update scoring software (McGrew, Schrank, & Woodcock, 2007).

Comparability of Methods

In previous studies of number-line estimation, students estimated silently and used pencil and paper or desktop computers. In contrast, in this study students made their estimates on a computer tablet while giving self-reports on two-thirds of the trials. As noted earlier, no differences were found in a comparison study using pencil-and-paper versus tablet methods for stimulus presentation with Grade 6 students (Appendix D). Whether asking students to give self-reports changed their performance on the task was addressed in a separate between-subjects comparison in Grade 6 students (Appendix B). No differences were found between Grade 6 students who estimated silently versus aloud.

To further explore whether self-reports affected children's thinking on number-line estimation, a 2(Condition: Silent, Aloud) x 2(Range: 100, 1,000) analysis of variance with repeated measures on both variables was conducted for each grade. Students in both Grades 2 and 4 had significantly greater error on 0-1000 than 0-100 range [Grade 2, $F(1,23) = 14.56$, $p = .001$, $\eta_p^2 = .39$, and Grade 4, $F(1,23) = 9.87$, $p = .005$, $\eta_p^2 = .30$], but there was no effect of self-report. Students in Grade 6 also showed greater error on 0-1000 than 0-100, $F(1,23) = 5.26$, $p = .03$, $\eta_p^2 = .19$, as expected, but there was also an interaction between condition and range, $F(1,23) = 5.27$, $p = .03$, $\eta_p^2 = .19$ ($p < .05$). In the silent condition Grade 6 students showed about the same mean percent of absolute error (PAE) for both 0-100 and 0-1000 (*M*s = 5%), whereas when reporting their thinking, Grade 6 students had less error for 0-100 ($M = 3\%$) than for 0-1000 ($M = 6\%$). This specific difference for Grade 6 students may have emerged because Grade 6 students showed less variability than Grades 2 and 4 students, which might have magnified the Condition x Range effects. Because all students estimated silently first, all students would have had the same advantage of silent practice before completing the self-report trials. Next explored was whether the data and patterns of error observed in this study are comparable to previous studies.

Comparing measures of linearity and accuracy. Traditionally, two analyses have been used to assess students' performance on number-line estimation: model fit and accuracy (Booth & Siegler, 2006; Opfer & Siegler, 2007; Siegler & Booth; 2004). To determine whether the current results in which students gave self-reports are consistent with past work, the conventional analyses of model fit and accuracy were explored on the 0-100 and 0-1000 ranges, for which there are comparisons in the literature. As in previous work, each estimate of a target was plotted as a function of the number's actual position. Perfect estimation would be fit by a linear model with a slope of 1.00 and an intercept of zero, that is, $y = x$ (Young & Opfer, 2011). Students' estimates, plotted as a function of the number's actual position are shown in Figure 5 for both 0-100 and 0-1000 across Grades 2, 4, and 6.

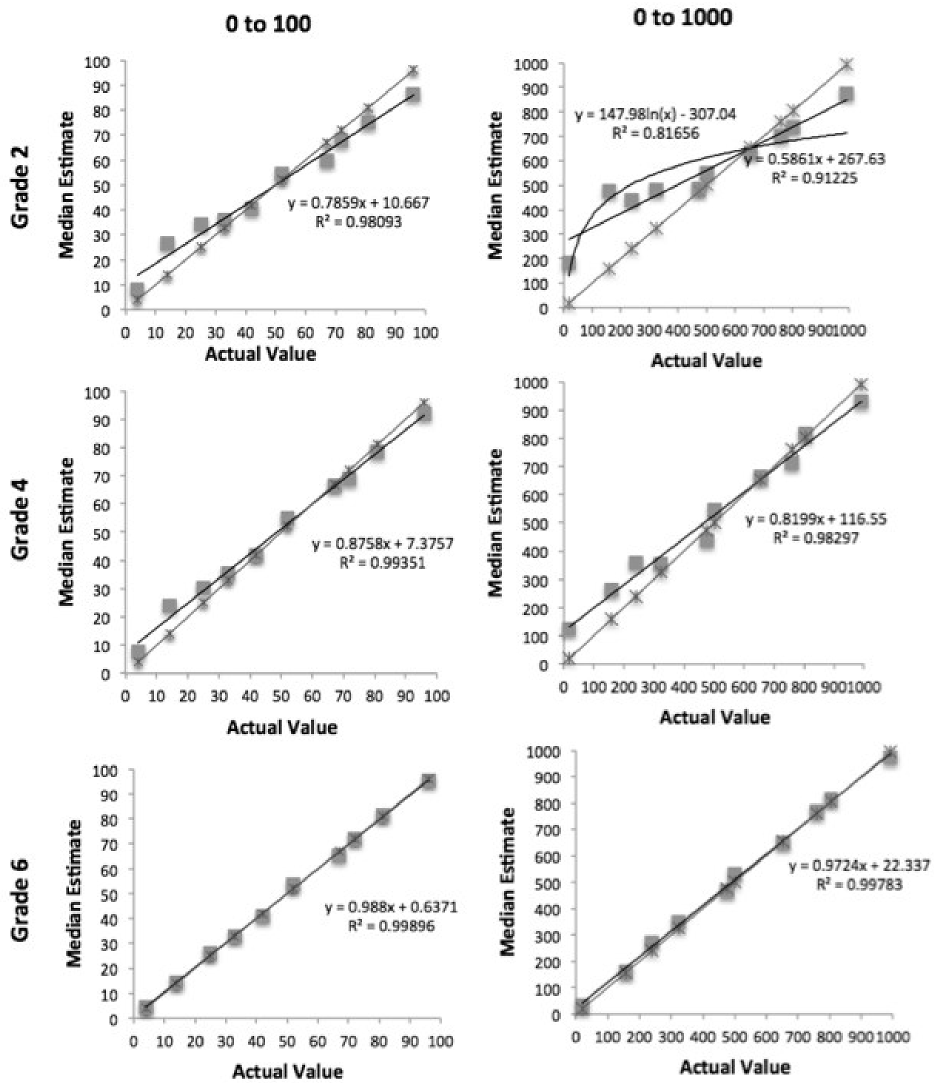


Figure 5. Best-fitting equations for both ranges by grade where the gray line illustrates perfect accuracy.

These results are consistent with past research, both in terms of shape and R^2 values for linear and logarithmic models (Table 4) (Booth & Siegler, 2006; Opfer & Siegler, 2007). For 0-100, students in Grades 2, 4, and 6 all showed fairly linear patterns of estimates. For 0-1000, estimates from Grade 4 and 6 were better fit by linear than logarithmic models. Grade 2 students showed a pattern of performance that was less linear on 0-1000 than 0-100, and less linear than students in Grades 4 and 6. These results in Grade 2 for 0-1000 are consistent with previous literature, although Grade 2 students in the present study show median estimates that are more linear than found in some previous studies (e.g., Opfer & Siegler, 2007; Thompson & Opfer, 2010). One explanation for this difference may be that in both previous studies, targets on 0-1000 were intentionally oversampled at the low end of the range to maximize the ability to discriminate between the logarithmic and linear models. In the current study, targets were evenly sampled across the 0-1000 line.

Table 4

Results of Current Study compared with Results from Previous Studies

Study	0-100						0-1,000					
	Grade						Grade					
	2		3 or 4		5 or 6		2		3 or 4		5 or 6	
	Log R^2	Lin R^2	Log R^2	Lin R^2	Log R^2	Lin R^2	Log R^2	Lin R^2	Log R^2	Lin R^2	Log R^2	Lin R^2
Ashcraft & Moore, 2012	.78[a]	.85	.82[a] (4th)	.95 (4th)	.79[a] (5th)	.96 (5th)			.76[a] (4th)	.87 (4th)	.82[a] (5th)	.92 (5th)
Booth & Siegler, 2006, Exp. 1	.88	.97	.85 (3rd)	.98 (3rd)								
Booth & Siegler, 2006, Exp. 2							.88	.91	.71 (4th)	.98 (4th)		
Opfer & Siegler, 2007, Exp. 1							.95			.99 (4th)		
Thompson & Opfer, 2010 Exp. 1							.91	.82	.87 (3rd)	.98 (3rd)	.84	1.00
Current Study	.86	.98	.84 (4th)	.99 (4th)	.83 (6th)	1.00 (6th)	.82	.91	.74 (4th)	.98 (4th)	.76 (6th)	1.00 (6th)

[a]Ashcraft and Moore found that exponential models fit their data better than logarithmic models in all but three cases (out of 124 students).

To assess how well students estimated, the percent absolute error (PAE) was calculated on each target to arrive at each student's mean PAE (Booth & Siegler, 2006). Previous research has shown that mean absolute error decreases with grade (Booth & Siegler, 2006; Siegler & Booth, 2004). Results from this study are consistent with past studies (Table 4). The expected pattern was found on both ranges. On the 0-100 number line, Grade 2 students' PAE, was slightly higher (9%) than those of students in Grades 4 and 6 (7% and 3%, respectively). The same pattern of grade differences was found for the 0-1000 line, where students in Grade 2 had higher error (PAE =17%) than students in Grades 4 and 6 (11% and 6%, respectively). In summary, analyses of model fits and accuracy indicate that, despite some methodological differences, this pattern of performance for Grades 2, 4, and 6 on 0-100 and 0-1000 lines was comparable to the results from other studies. With comparability to previous studies established, the focus shifts to characterizing procedural number-line knowledge, and explicit conceptual number-line knowledge along with examining the relations among procedural knowledge and implicit and explicit conceptual knowledge on the number-line task.

Explicit Number-line Knowledge

Nine items captured children's explicit conceptual knowledge about numerical categories (general location of small, medium, and large numbers), equal intervals on 0-100, whether and how the locations of 50 and 60 changed when the scale changes to 0-1000, as well the locations of proportions on 0-100 (Appendix C). Summing across the nine items yielded a K_{num} score for each student. Scores ranged from 1 to 9 with a mean of 5.45 (SD = 2.04).

Two questions about children's explicit conceptual knowledge of the number line were addressed. First, in quantitative terms, how does K_{num} change with grade? Because K_{num} was designed to capture a range of explicit knowledge about number-line estimation including basic concepts such as numerical categories and more advanced concepts such as proportion, K_{num} was predicted to increase with grade. A one-way ANOVA confirmed the prediction (Ms = 3.96, 5.32, and 7.08 for Grades 2, 4, and 6, respectively), $F(2,69) = 22.71$, $p < .01$, $\eta_p^2 = .40$. One concern is that younger students may have misunderstood the questions and language used on the explicit number-line knowledge task. Children, however, are able to produce the kinds of sentence constructions used on the task, such as *wh-* questions and *if-then* statements, by age 4, and receptive understanding generally precedes language production (Rutherford, 1993). Therefore, language on the task was appropriate for even the youngest children in this study.

Second, in more qualitative terms, how does K_{num} change with grade? Figure 6 illustrates the extent to which students in each grade correctly understood each K_{num} item. The items in Figure 6 are ordered to reflect how many students across the three grades answered each item correctly. Figure 6 illustrates clear differences in the grade at which students are able to explicitly demonstrate their knowledge of various number-line concepts. For example, most students in Grades 4 and 6 correctly answered the two items probing whether they understood that intervals between numbers are constant across the line. In contrast, fewer than half of the Grade 2 students seemed to understand the concept of equal intervals across the line. Students' explicit and robust understanding of the concept of half on the number line seemed to emerge with increasing grade. Previous research has shown that with grade, students improve in their ability to estimate, compute and use fractions (Hecht & Vagi, 2010). Moreover, the clear emergence of half on the number line is consistent with the role of "half" as a crucial boundary in learning about proportions (Spinillo & Bryant, 1991). All Grade 6 students explicitly demonstrated an understanding of half on the number line; in contrast a third of the Grade 4 students and nearly half of the Grade 2 students missed this item.

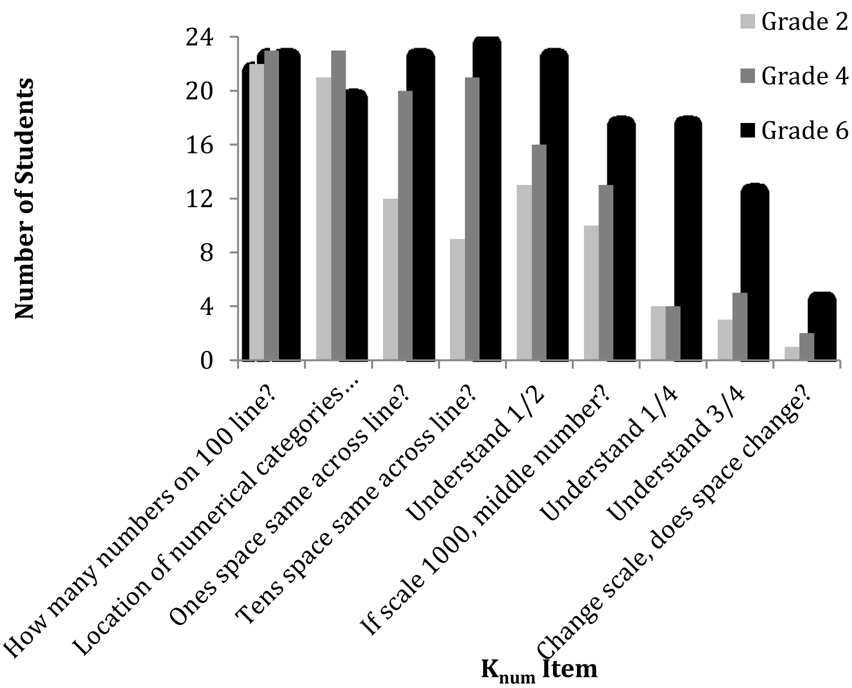


Figure 6. Number of students per grade (maximum of 24) answering each K_{num} item correctly.

The last four items in Figure 6 clarify further what kinds of explicit concepts about number lines might emerge with age and more experience. Most students in Grade 6 understood that if the scale changed from 100 to 1,000, then the middle number would change from 50 to 500, compared with only about half the students in Grades 2 and 4. The proportions of ¼ and ¾ were clearly more advanced conceptual properties of the number line given that both of these were only understood by a handful of students in Grades 2 and 4. Because more Grade 6 students understood ¼ than ¾ it may be that explicit knowledge of ¼ emerges before ¾ but this interpretation would have to be explored further. Finally, the most difficult item, understanding that if the scale changes from 100 to 1,000 then the numbers 50 and 60 would both move down on the number line and be closer together, was answered correctly by only about 11% of all students with most being in Grade 6. These results demonstrate that explicit conceptual knowledge of the number line changes with age, as expected, and that some conceptual ideas about the number line are more advanced than others. What remains to be seen is whether, and to what extent, K_{num} is related to the measures of implicit conceptual and procedural knowledge.

Relations among conceptual measures of number-line knowledge. First, does students' explicit conceptual knowledge about the number line, K_{num}, correlate with the implicit conceptual measures of number-line knowledge, mean PAE and linearity of the overall pattern? These two measures were highly and negatively correlated (*r*s > -.85) and so the strength of their relations with K_{num} was expected to be similar. A partial regression analysis, controlling for grade, confirms that higher K_{num} scores were associated with less estimation error for both ranges, [$r(69) = -.43$ for 0-100 and $r(69) = -.49$ for 0-1000, *p*s <.01, and for higher R^2_{Lin} for both ranges ($r(69) = .38$ for 0-100 and $r(69) = .34$ for 0-1000, *p*s <.01]. The strength of the relation between K_{num} scores and both measures of implicit conceptual knowledge varied, however, when correlations were explored separately for each grade and range (Table 5).

Table 5

Correlations Between K_{num} and Measures of Implicit Conceptual Number-Line Knowledge.

Implicit Conceptual Measure	0-100 Grade 2	0-100 Grade 4	0-100 Grade 6	0-1000 Grade 2	0-1000 Grade 4	0-1000 Grade 6
R^2_{Lin}	.44*	.43*	.16	.24	.52*	.72**
Mean PAE	-.46*	-.51*	-.22	-.29	-.77**	-.58**

*$p < .05$, ** $p < .01$, $df = 22$.

Differences in the strength of the relation demonstrate that the explicit conceptual knowledge captured in K_{num} does not contribute to students' errors in the same way across grades for both ranges. K_{num} was related to performance on the 0-100 line for students in Grades 2 and 4 but not in Grade 6. Recall that on the 0-100 number line, Grade 2 and 4 students' PAE was slightly higher (*Ms* = 9% and 7%, respectively, *SD* = 4% for both) than the PAE of students in Grade 6 (*M*= 3%, *SD* = 2%). Not finding a strong correlation for Grade 6 students may be the result of Grade 6 students having performed so well on 0-100 that the amount of variability to be explained is too small. On the 0-1000 line, K_{num} was not related to estimation performance for students in Grade 2, but was related as predicted for students in Grades 4 and 6. K_{num} scores for Grade 2 may not have predicted performance on 0-1000 because the number-line knowledge task was designed, generally, around the 0-100 line and only some of the more difficult questions, about scale change, for example, allowed students to demonstrate knowledge of the 0-1000 line. As a result, the explicit number-line knowledge task may have not been sensitive to measuring what Grade 2 students might know about 0-1000 and therefore not predicative of their accuracy in estimating on 0-1000. Whether K_{num}, as a measure of explicit conceptual number-line knowledge, will shed more light on the development of students' procedural knowledge is explored later in this study.

Analysis of Students' Number-line Estimation Processes, Procedures, and Tactics

Students' self-reports and observed behaviours were examined to investigate the ways students' solution operations for number-line estimation may change with grade and range. The analyses proceeded in several steps. First, students' self-reports on each trial were coded based on the proposed task analysis (Figure 3) at the first level of problem-solving operations, processes (Table 1). Second, patterns of processes were identified and classified into the next level of problem-solving operations, procedures. Third, I investigated whether and how students' selection of procedures varied, both qualitatively and quantitatively, with increasing age and for 0-100 and 0-1000 number lines. Fourth, to characterize how students selected procedures as a function of target within each range, I investigated the next level of solution-operations, tactics.

Fifth, students' use of tactics was explored, both qualitatively and quantitatively, with increasing age and range. Sixth and finally, I explored whether and how students' adjusted their tactics as they moved from 0-100 to 0-1000.

Processes. Relations among processes and how students combined processes formed the basis of problem-solving procedures. Because procedures can involve several processes in various combinations and patterns, children's self-reports were coded at the level of processes. Characterization of how students estimated was at the level of procedures. The coding scheme reflects the general processes described in the rational task analysis in Figure 3 (see Appendix E for coding protocol). Twenty-two specific processes were identified and coded: two specific reformulation processes (Table 6), ten specific anchoring processes (Table 7), nine specific adjustment processes (Table 8), and guessing. Guessing was the absence of using any anchors or adjustment and was characteristically accompanied by the child saying "guessed" or "don't know". Two additional pieces of information were coded for each trial: (a) the kind of evidence (gesture and speech) the coder relied on in making her judgement, and (b) whether the child changed his or her answer or abruptly shifted to a different process mid-estimation (Appendix E). The two experimenters who administered the task coded the data by watching videos and using the coding scheme and principles outlined in Appendix E. Inter-rater reliability was 96% on a sample of data from 12 participants (four per grade).

Table 6

Coding Scheme for Number-line Estimation Process: Reformulation

Specific Process	Description	Characteristic Responses
Target	By truncation, rounding or changing form	Target is 52: "looking for 50"
Scale	By truncation, rounding or changing form	On 0-100: "think about as 0 to 10"

Table 7

Coding Scheme for Number-line Estimation Process: Anchoring

Specific Process	Description	Characteristic Responses
0 End	Refer to 0 endpoint	"here is 0"
High End	Refer to high endpoint, e.g., 1000 or 531	"here is 1000"
Middle	Refer to the middle, e.g., high endpoint/2, ½, halfway, or physical middle even if they clearly assign wrong number to the middle	"half is here"
Quarter	Refer to ¼ or ¾ or demonstrate splitting line into fourths	"25, 50, 75"; "three-quarters here"
Other Proportion	Refer to any other explicit proportion	"this is about a fifth"; "33, 66"
Previous Trial	Refer to an anchor from a previous trial	"45 was here last time"
Counted	Construct anchor by counting at least two sequential numbers	"10, 20, 30"
Construct Proportion	Construct anchor by constructing at least one proportion or counting several proportions between 2 anchors	"middle of the middle"; "0 here and 100 here, so one third, two thirds"
Non-proportional Construction	Construct ad hoc anchor from incremental or decremented move such as adding or subtracting exact values, moving up or down either in vague terms ("up a little") or in precise measured terms ("up 2 cm"), or finding an anchor between 2 other anchors	" add 10 on so here is 20"
Other	Other kinds of ad hoc anchors, including vague references to an area	"in the 30s" or "41 here"

Table 8

Coding Scheme for Number-line Estimation Process: Adjusting

Specific Process	Description	Characteristic Responses
Qualitative	Adjust qualitatively	“up/down”; “a bit”; “some”; “near”; “far”; “close to”; “far from”
Quantitative	Adjust quantitatively	“add/subtract”; “about 3 up”; “1 mm”; “greater or less than”, “a few ticks over”
Count by ones	Adjust by counting two or more sequential numbers by ones	“up 1,2,3,4”; “71,72,73,74,75,76”
Count by tens	Adjust by counting two or more sequential numbers by tens	“60, 70”; “510,520,530”
Count other	Adjust by counting two or more sequential numbers by some unit other than 1 or 10	“3, 6, 9”; “210,212,214,216”
Middle	Adjust by finding “middle” or “halfway between X & Y”	“25 is halfway between 0 and 50” or “go between 500 and 600”
Proportional	Adjust proportionally	“quarter of the way from 100”, “third of way from 10”
Previous Trial	Adjust using amount from a previous trial	“go about 10 like last time”
Guessing	Adjust by guessing	“then I guessed”, “somewhere around here”, “maybe here”, “around”, a back-and-forth gesture

The coding scheme for processes in number-line estimation captured the scope and richness of how students think about the number-line task, on both typical and atypical number-line ranges. Across both typical and atypical ranges, self-report data were collected on 2,880 trials (40 trials per student x 24 students x 3 grades). Of those, only 11 trials contained too little information to code (see Appendix E for details). On another 3% of trials (15 of remaining 1,437 typical trials; 28 of remaining 1,432 atypical trials), students changed their solution processes part way through their estimations. Though trials on which students changed their solution processes were successfully coded, they were removed from subsequent analyses because they are qualitatively different from trials where students do not interrupt or change their processes. Of the remaining trials nearly all could be coded using the scheme derived from the task analysis. Less than 2% of the trials (25 of remaining 1,437 typical trials; 17 of remaining 1,432 atypical trials) did not fit the coding scheme and were categorized as *other* processes (for details on the nature of *other* processes see Appendix E).

To illustrate how the coding scheme captured the range of students' responses, consider several example responses from the data. A Grade 2 student, when estimating 42 on a 0-100 number line, said, "50 is about there *(gesturing to the middle)*, and 42 is about there *(moving her hand lower from 50)*, and close to 50." The student's verbal explanation and gesture were used to code her response as anchoring to the middle (Table 7) followed by a qualitative adjustment (Table 8). A Grade 6 student estimated 805 on a 0-1000 number line and said "805, um, split into 200s (as he pointed to four 200 marks along the line) and up .5 centimeters." His response was coded as anchoring by constructing proportions (Table 7) followed by a quantitative adjustment (Table 8). A final example is from a Grade 6 student who estimated 158 on 0-1000 and said, "10 on a 100 scale, so close to 0." In this case, the student reformulated both the target (158 to 10) and the scale (1000 to 100) (Table 6), anchored to 0, and used a qualitative adjustment, "close to 0" (Tables 7 and 8, respectively).

Establishing the validity of the coding scheme and task analysis of number-line estimation is important because subsequent analyses of higher levels of problem-solving operations were based on organizing solution processes into procedures and, later, tactics.

The coding scheme of number-line estimation processes captured, in large part, the range of students' answers, which lends face validity to the task analysis and coding scheme. Assessing the external validity of the coding scheme and task analysis is difficult because there are few comparable measures of number-line estimation processes. Schneider and colleagues' (2008) data on eye fixations may reflect most closely the level of processes found in these data. Schneider et al. found that students in Grades 1 through 3 looked more at the end points and middle of the line when estimating on a 0-100 line. This pattern of looking, as well as converging results from two behavioural studies using accuracy and response latencies (Ashcraft & Moore, 2012; White & Szűcs, 2012), lend support for anchoring to points such as ends and middle when estimating. The coding scheme used here captures those, and other, anchoring processes.

Finally, though not presented as processes, earlier studies of number-line estimation (Newman & Berger, 1984; Petitto, 1990) included all of the processes introduced here including anchoring to the endpoints and middle, adjusting by counting, and guessing. Petitto (1990) noted, "Children…sometimes used combinations of strategy components on a single problem" (p. 67). Earlier pilot studies of this work also revealed that students used a combination of processes to estimate. Because processes are used in combination to estimate a target, the combinations of processes, or procedures, students used to estimate are investigated next. Because in-depth analyses of the atypical ranges were beyond the scope of this study, from here forward I focused on the two typical ranges: 0-100 and 0-1000.

Procedures. At this juncture the research question that emerges is: What procedures do students use to solve number-line estimation problems? To answer that question, students' number-line estimation procedures were classified by organizing their number-line processes into 41 specific procedures that were collapsed into eight general procedures. Briefly, from the data on processes, eight general patterns of anchoring processes and five general adjustments were identified, revealing 40 possible specific procedures (see Appendix F for details). *Guessing* was included as the forty-first specific procedure. To illustrate: To estimate 52 on 0-100 a student might anchor to the middle, and then construct an adjustment, “up a little bit”. This would be the specific procedure: *construct* [adjustment] *from middle* [anchor]. To estimate 4 on 0-100 a student might anchor to 0 and then count up by ones to 4. This would be the specific procedure: *count* [adjustment] *from 0* [anchor]. A final example is to estimate 19 on 0-100, a student might anchor to “the 20s”, therein constructing an *ad hoc* anchor, and go “down 1”. This would be the specific procedure: *construct* [adjustment] *from ad hoc* [anchor]. Of the 41 possible specific procedures identified, students used 38 specific procedures at least once, thus showing a remarkable amount of variability in how they estimated targets on the number line, variability not reflected elsewhere in the literature.

Of the 38 specific procedures, some were used rarely: by only one or two students on one or two targets. Fully 96% of the estimates, however, were captured by 24 specific procedures (Table F2). Students used three specific procedures on between 16 and 18% of trials, which, taken together, accounted for how students estimated on 50% of the trials. The three most frequently used specific procedures entailed constructing an adjustment from either end or from the middle. Identifying the three most frequently used specific procedures is important because they are similar to the kinds of procedures observed in earlier work (Newman & Berger, 1984; Petitto, 1990). Strikingly, 21 additional specific procedures accounted for the remaining 46% of the trials. This novel finding underscores the remarkable variability in how students estimate and the usefulness of having taken a process-driven approach to cataloguing the ways students estimated. To preserve the observed variability while still making it possible to consider patterns in how procedures were selected for different targets, the 38 specific procedures were organized into eight general procedures.

Two constraints guided the organization of the 38 specific procedures into eight general procedures. First, in most cases, specific procedures were organized with reference to their underlying anchoring pattern in relation to the number line including: *0 end*, *high end*, *middle*, *ad hoc*, *segments*, and *proportional* anchoring patterns. For example, the specific procedure *count from high end* was grouped into a general procedure, *High End*, along with the specific procedures *construct from high end*, *guess from high end,* and *high end only*. The one exception to using anchoring patterns to classify general procedures was when a student used a proportional adjustment within a segment such as the whole line or upper or lower half of the line. Using a proportional adjustment within a segment was classified as the general procedure *Proportional Segment*. For example, estimating 25 by anchoring to 0-end and the middle and proportionally adjusting, "halfway between the two" was coded as a Proportional Segment general procedure. The specific procedure *guess* remained its own general procedure, *Guess*.

In anticipation of classifying the selection of procedures as a function of target, an additional classification constraint was applied to the general procedures: specifying whether adjusting from an end was appropriate. To evaluate the appropriateness of adjusting from the wrong or right end as Petitto (1990) did, the use of 0 End or High End general procedures were reclassified based on the proximity of the end relative to the target. When an end was used for a target on the close side of half (50 or 500), the general procedure was called *Ends*. When an end was used for a target on the far side of half (50 or 500), the general procedure was reclassified as *Wrong End*. For example, the appropriate use of 0 End to estimate 4 was classified as Ends, as was the appropriate use of the High End to estimate 96. In contrast, the inappropriate use of 0 End to estimate 72 was classified as Wrong End, as was the inappropriate use of High End to estimate 17. All other general procedures remained the same. The final set of eight general procedures, along with the number of students per grade who used each general procedure, is shown in Table 9. General procedures were organized into four levels of sophistication from immature to advanced based on the assumption that students' use of procedures may reflect the use of less number-line knowledge, as in guessing, to more advanced and integrated number-line knowledge, as in anchoring to proportions.

Table 9

Number of Students per Grade Using a General Procedure at Least Once in a Range

		Range					
		0-100			0-1000		
		Grade			Grade		
Level	General Procedures	2	4	6	2	4	6
Immature	Guess	3	0	0	4	2	1
	Wrong End	6	1	0	11	6	2
Basic	Ends	22	23	23	24	23	23
	Middle	18	20	23	15	22	24
Intermediate	Ad Hoc	13	9	8	17	15	8
	Segments	18	16	13	11	13	13
Advanced	Proportional Segment	3	3	13	2	5	12
	Proportions	1	2	10	0	2	5

Use. Having identified students' general procedures, explored next is whether the use of general procedures might reasonably be ordered from immature to advanced as shown in Table 9. I examined whether and how mean proportion use of each level of general procedures changed with grade and range. For example, does the mean proportion use of more advanced general procedures increase with grade as predicted, and decrease in use from 0-100 to 0-1000 as expected? Table 9 shows the number of students per grade who used each general procedure in each range, and Table 10 reveals the number of students who used each general procedure at least once across *both* ranges.

Table 10

Number of Students per Grade Using a General Procedure at Least Once in Both Ranges

		Grade		
Level	General Procedure	2	4	6
Immature	Guess	2	0	0
	Wrong End	5	0	0
Basic	Ends	22	22	23
	Middle	13	19	23
Intermediate	Ad Hoc	9	7	6
	Segments	8	10	8
Advanced	Proportional Segment	1	1	6
	Proportions	0	1	4

Table 11 shows the mean proportion use (trials used/total trials) for each general procedure by grade and range. Three trends stand out across Tables 9, 10, and 11. First, nearly all students used ends-based general procedures in both ranges. Second, the number of students using procedures based on the middle in both ranges nearly doubles from Grade 2 to 6. Third, the two most immature and most advanced procedures were used by the fewest students and showed the expected patterns where older students used immature procedures less and advanced procedures more than younger students.

Table 11

Mean Proportion Use of General Procedures by Range and Grade

		Range					
		0-100			0-1000		
		Grade			Grade		
Level	General Procedure	2	4	6	2	4	6
Immature	Guess	.01	.00	.00	.04	.03	.01
	Wrong End	.03*	.02	.00	.09*	.04*	.01
Basic	Ends	.46*	.45*	.29*	.35*	.41*	.34*
	Middle	.16*	.27*	.31*	.17*	.25*	.30*
Intermediate	Ad Hoc	.18*	.11*	.07*	.22*	.13*	.06*
	Segments	.13*	.14*	.13*	.12*	.10*	.15*
Advanced	Proportional Segment	.03	.01	.06*	.01	.03*	.08*
	Proportions	<.01	.01	.14*	.00	.01	.06

*95% confidence interval does not include 0.

Because few students used immature and advanced procedures across both ranges (Table 10), ANOVA was an inappropriate inferential test for those procedures. For procedures in the basic and intermediate levels, 3(Grade) x 2(Range) ANOVAs with repeated measures on the last variable were conducted for each general procedure when more than a quarter of students used the general procedure in both ranges (Table 10). When appropriate, contrasts were conducted to reveal the nature of grade effects (Hale, 1977).

Immature procedures. Because immature procedures, Guessing and Wrong End, were used rarely and by only a few students across both ranges (Table 10), analysis is descriptive. As expected, students in Grade 2 used immature procedures more than students in Grades 4 and 6 (Table 9). When a student used an immature procedure in both ranges (Table 10), it was always a younger student. The use of Guessing or Wrong End appears to increase from the 0-100 to 0-1000 both in terms of number of students (Table 9), and mean proportion of use (Table 11). This pattern may indicate that as students face a less familiar or more difficult range, they recruit less advanced procedures.

Basic procedures. The two basic general procedures, Ends and Middle, were frequently used across all grades and both ranges. The use of basic procedures was expected to decrease with grade as students replaced basic with more advanced procedures. Table 9 illustrates that nearly all students in all grades used Ends at least once. Grade 2 and 4 students used Ends (M = .41, M = .43, respectively) more than Grade 6 students (M = .31), $F(2,69) = 3.24$, $p < .05$, $\eta_p^2 = .09$. Use of Ends was not affected by range, but grade interacted with range, $F(2,65) = 4.30$, $p < .05$, $\eta_p^2 = .11$, because both Grade 2 and 4 students used Ends more on the 0-100 line than on the 0-1000 line whereas Grade 6 students used Ends slightly more on the 0-1000 line than the 0-100 one (see Table 11). Grade 6 students may have used Ends less on the 0-100 line compared to younger students because they may have at their disposal, especially on a range they presumably had more experience with, a wider range of general procedures to select from such as anchoring near proportions like quarter and three-quarter. Middle was used by at least half the students in each grade across both ranges (Table 10). In contrast to Ends, use of Middle increased as a function of grade (Ms = .17, .26, and .31 for Grades 2, 4, and 6, respectively), $F(2,69) = 7.39$, $p < .01$, $\eta_p^2 = .18$, but was unrelated to range.

Taken together, these results are consistent with the idea that students may initially orient to the two ends, as noted in eye tracking studies (Schneider et al., 2008; Sullivan et al., 2011), but with age and experience they increasingly orient to the middle (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Newman & Berger, 1984; Petitto, 1990; Rouder & Geary, 2014; Siegler & Opfer, 2003; White & Szűcs, 2012). Overall this pattern of shifting reliance on different procedures as a function of grade is consistent with the idea that procedures seem to emerge in a sequence. The frequent use of Ends and the emerging use of Middle highlights the extent to which Ends and Middle may be basic and foundational procedures for number-line estimation.

Finally, despite presumable increases in knowledge and experience, the use of basic procedures across all grades illustrates that these procedures will always be appropriate for some targets. For example, even if a student knows about quarters and halves, if asked to estimate 3 on a 0-100 line, it still makes sense to anchor to the 0 end and adjust up rather than anchor to 1/4, or 25, and adjust down “about 20”.

Intermediate procedures. Next examined were Ad Hoc procedures. Though grouped into the intermediate level because they may reflect increasing differentiation along the line, Ad Hoc procedures were not as well characterized as the other general procedures (see Appendix F for details) and may represent a grab bag of thinking from “educated guessing” Ad Hoc to “thoughtfully constructed” Ad Hoc procedures. Use of Ad Hoc procedures decreased with grade (*Ms* = .20, .12, and .07, for Grades 2, 4, and 6, respectively), $F(2,69) = 4.04$, $p < .05$, $\eta_p^2 = .11$, but was unrelated to range. The use of Segments did not vary as a function of either grade or range. Table 10 shows that about the same number of students in all three grades used Segments in both ranges.

Overall, classifying Ad Hoc and Segments as intermediate procedures may be inappropriate for a few reasons. First, all Ad Hoc procedures are constructed procedures but are not all necessarily constructed in the same way. That is, individual children may construct and devise Ad Hoc procedures in idiosyncratic ways that might reflect differing levels of differentiating across the line. Second, although using Segments of the line seems intuitively more complex than anchoring in a single point because the student is anchoring to a segment of the line such as the lower or upper half, the evidence shows that students in all grades used Segments to the same extent. It is not clear that either Ad Hoc or Segments are really intermediate procedures. Segments might better be thought of as variants of basic procedures like Middle and Ends, and Ad Hoc procedures may be too varied to classify.

Advanced procedures. Because few students used Proportional Segments and Proportions across both ranges (Table 10), analysis is descriptive. As expected, older students appear to have used advanced procedures more than younger students (Tables 9 and 10). Because students may have more advanced procedures at their disposal on a more familiar range, students were expected to use advanced procedures more on the 0-100 than the 0-1000 line. This expected pattern was found for Proportions, but not for Proportional Segments. These results suggest that advanced procedures are associated with higher grades and potentially with more knowledge about the number line but, overall, use was too low to be certain. The infrequency of advanced procedures may be due to few appropriate opportunities to use them. For example on 0-100, only the targets 25, 33, and 72, easily afforded the opportunity to use Proportional Segments or Proportions. Had more targets around common proportions been included, like ¼, ¾, 1/3, or 2/3, there may have been more use of advanced procedures.

Summary. Overall the evidence on use of general procedures shows some support for the proposed ordering of procedures from immature to advanced. Specifically, at the extremes, immature procedures were used more by younger students and advanced procedures more by older students. The two intermediate procedures did not clearly fall into the intermediate class. Segments might better fit the basic level because they entail using Ends and Middle, albeit in tandem rather than alone. Ad Hoc procedures might require further characterization to establish the level of sophistication in differentiating along the line to tease apart the "educated guesses" from the "thoughtfully constructed" procedures. The frequent use of basic procedures strengthens the idea of some procedures being foundational. Further, the emergence, with age, of procedures involving anchoring to the middle is consistent with the idea that procedures may emerge sequentially.

Relation between procedures and accuracy. Next, if procedures are ordered to reflect increased differentiation across the line, more advanced procedures should result in more accurate estimates. Arguably, however, assessing accuracy as a function of target alone fails to account for the within-task variability inherent in number-line estimation. Each target has unique affordances by virtue of its location along the number line (Ashcraft & Moore, 2012; White & Szűcs, 2012). For example, a target near the middle is likely to be approached differently than a target near 0. Figure 7 shows that different procedures are selected for different targets illustrating how using a particular procedure results in less error for some targets and higher error for others. For example, using Middle for targets near the middle results in lower error than using Middle for targets near the ends of the line (Figure 7). To characterize the within-task variability in accuracy unique to each student, I investigated tactics, that is, how students selected procedures within a range as a function of target.

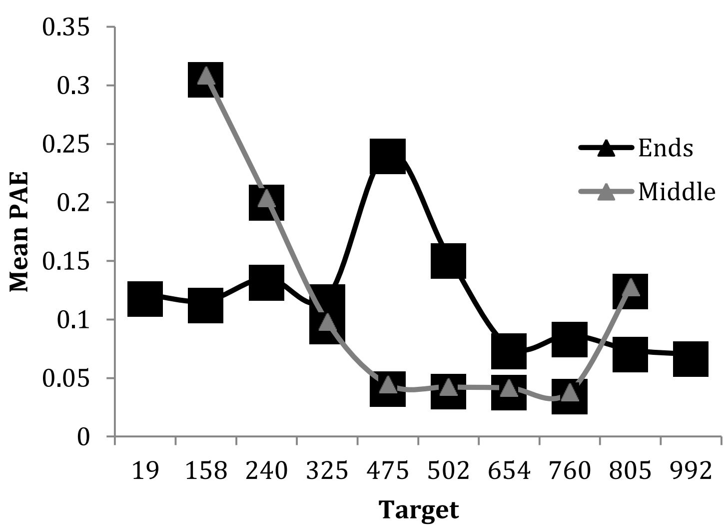


Figure 7. Mean PAE by target on 0-1000 for two general procedures collapsed across grade.

Tactics. Investigating students' use of general procedures without accounting for a target's position along the line is problematic because different targets afford the use of different procedures (Newman & Berger, 1984; Petitto, 1990; White & Szűcs, 2012). For example, in Figure 8, across all grades, students clearly used the two basic general procedures, Ends and Middle, on different targets. Specifically, and converging with Ashcraft and Moore (2012) and White and Szűcs (2012), Ends was used more for targets near the ends and Middle was used for targets near the middle. In contrast, Figure 9 shows a different relation between targets and advanced general procedures, Proportional Segments and Proportions. In contrast to younger students, Grade 6 students used advanced general procedures more and did so for targets *between* the ends and middle of the line in a M-shaped pattern confirming the use of certain procedures as a function of target as suggested in Ashcraft and Moore's (2012) error analysis as a function of target. In the current study, nearly half the Grade 6 students used Proportional Segments when the target was 25 on 0-100. This same pattern appears to be only just emerging for younger students. For Grade 6 on 0-100, Proportions were also used for the spaces between the ends and middle: around 25 and 33, and 72 and 81. Collectively these results illustrate how different targets afford the use of different general procedures and the need to identify individuals' tactics within each range.

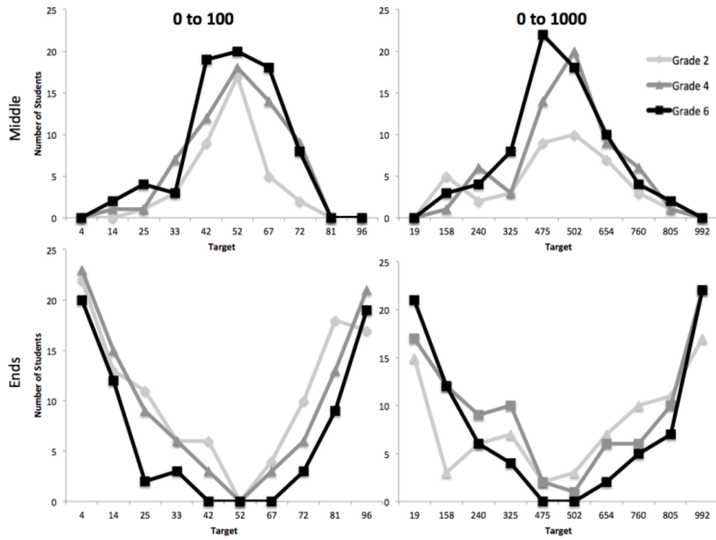


Figure 8. Use of basic general procedures by grade as a function of target.

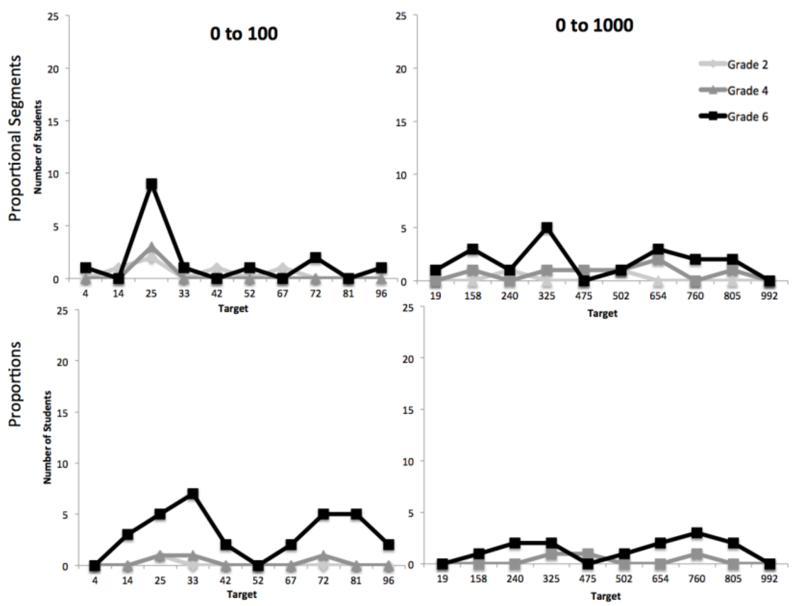
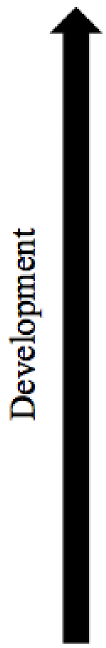


Figure 9. Use of advanced general procedures by grade as a function of target.

Classification. Initially, for each of 72 students, I examined the general procedures students selected across all 10 targets on the 0-100 number line. Though there were similarities and clear patterns, no two children selected the same general procedure for all 10 targets, highlighting once more, the incredible variability in how students assemble solution procedures within a seemingly highly constrained task space. Because there are more than a billion possible target by procedure permutations, or tactics, (8 procedures by 10 targets, or 8^{10}), expectations about the kinds of tactics to be identified were initially drawn from the proposed sequence of tactics in the tactical ladder (Figure 3).

When initially classifying tactics using the proposed tactical ladder, I noticed students used two tactics not originally proposed: (a) Ad Hoc procedures in several quintiles of the line, and (b) Proportional Segments on at least one of three middle quintiles of the line in a way not dissimilar to the idea of judging part of a whole (Barth & Paladino, 2011; Slusser et al., 2013). The proposed tactical ladder was revised to include *Ad Hoc* and *Proportional Segment* tactics (Figure 10). A tractable coding scheme, shown in Table 12, was developed based on the identified tactical ladder (Figure 10). Table 12 highlights the criteria used to judge each student's pattern of general-procedure use in relation to the quintiles of the line (for details, see Appendix G). For example, on the 0-1000 number line, a Grade 2 student used Ends on six targets and Wrong End on four targets. Her tactic was classified as *Ends Only* (Table 12). A Grade 6 student used the Ends general procedure appropriately in the first and last quintiles as well as for one target each in the second and fourth quintiles. For the remaining targets, he used Middle. His pattern of procedures was classified as the *Landmark Middle* tactic. Using the coding scheme in Table 12, for both 0-100 and 0-1000 ranges, less than 1% of all data were deleted (one student was missing data on more than 40% of trials on 0-100). Only eight students were coded as using an *Other* tactic on at least one range (~5% of data) (Table 13). The tactics of the remaining 64 students were successfully coded for both ranges.



<table>
<tr><th rowspan="2">Skill Level</th><th rowspan="2">Tactic</th><th colspan="5">Quintile on Number Line</th></tr>
<tr><th>Origin</th><th>First Quarter</th><th>Middle</th><th>Third Quarter</th><th>End</th></tr>
<tr><td>Expert</td><td>Pure Proportions</td><td colspan="5">Proportions</td></tr>
<tr><td>Advanced</td><td>Proportional Landmarks</td><td>Ends</td><td>Proportions</td><td>Middle</td><td>Proportions</td><td>Ends</td></tr>
<tr><td>Intermediate</td><td>Proportional Segments</td><td>Ends</td><td colspan="3">Proportional Segments or Middle</td><td>Ends</td></tr>
<tr><td>Basic</td><td>Landmark Middle</td><td>Ends</td><td colspan="3">Middle</td><td>Ends</td></tr>
<tr><td rowspan="2">Emerging</td><td>Ad Hoc</td><td>Ends</td><td colspan="3">Ad Hoc</td><td>Ends</td></tr>
<tr><td>Ends Only</td><td colspan="5">Ends</td></tr>
<tr><td>Immature</td><td>Guessing</td><td colspan="5">Guess</td></tr>
</table>

Figure 10. Identified number-line estimation tactics.

Table 12

Coding Scheme for Number-line Tactics

Level	Tactic	Inclusionary Criteria	Exclusionary Criteria
VII	Pure Proportions*	Proportions in every quintile	
VI	Proportional Landmarks*	Proportions at least once	Guess
V	Proportional Segmenting*	Proportional Segments at least once	Proportions and Guess
IV	Landmark Middle*	Ends or higher at least once in each end quintile; Middle or Segments in middle quintile; **May include:** *Ad Hoc*	
III	Ad Hoc*	Ends at least once in each end quintile; Ad Hoc in each of three middle quintiles and at least 40% of trials overall within a range	Middle used in middle quintile
II	Ends Only	Ends in each of 5 quintiles; at least 50% of trials total; **May Include**: Several *Wrong End*; *Guess* on no more than 1 trial	Middle used in middle quintile
I	Guessing	Guessing on at least 30 % of trials overall across all quintiles within a range	Proportional Segments or Proportions
O	Other	Does not fit any pattern above; notes in Appendix G	
X	Deleted	Missing at least 40% of trials per range	

Note. Inclusionary and exclusionary criteria refer to obligatory features, unless noted as "may include."

*Denotes tactics that include features of one level lower.

Table 13

Mean Proportion of Students per Grade (n = 24) Using Number-line Tactics by Range

Tactic	0-100			0-1000		
	2	4	6	2	4	6
Pure Proportions	0.00	0.00	0.00	0.00	0.00	0.00
Proportional Landmarks	0.04	0.08	0.42	0.00	0.08	0.21
Proportional Segmenting	0.13	0.13	0.42	0.08	0.25	0.42
Landmark Middle	0.58	0.63	0.17	0.42	0.46	0.38
Ad Hoc	0.08	0.04	0.00	0.21	0.00	0.00
Ends Only	0.08	0.04	0.00	0.17	0.00	0.00
Guessing	0.00	0.00	0.00	0.04	0.08	0.00
Other	0.08	0.04	0.00	0.08	0.13	0.00
Deleted	0.00	0.04	0.00	0.00	0.00	0.00

Because it is unclear precisely what students using the Other tactic were doing, these students were excluded from subsequent analyses, leaving 64 students in total (20 in each of Grades 2 and 4, and 24 in Grade 6). Also, no student used a Pure Proportion tactic so it is not considered further. The numbers of students per grade using each of the six remaining tactics on the 0-100 and 0-1000 number line are shown in Table 14. In total, excluding tactics that were deleted or coded as Other, 94% of the tactics of all students were successfully coded. The tactics of both number lines were coded for 89% of students. To the best of my knowledge, this study is the first to identify the unique tactics at this level of detail that individual students used for a range on number-line estimation. These results strengthen the claim that students approach ranges tactically as suggested in the literature (e.g., Ashcraft & Moore, 2012; Barth & Paladino, 2011; Newman & Berger, 1984; Siegler & Opfer, 2003; Petitto, 1990; Siegler & Booth, 2005; Slusser et al., 2013). In addition, some students used tactics not previously suggested in the literature such as the Ad Hoc tactic. Of note is that the tactics have been organized to represent climbing from reliance on tactics requiring less knowledge about the number line, like guessing, to tactics that require students to increasingly differentiate and hierarchically integrate their number-line knowledge such as in the use of a Proportional Landmark tactic. To explore whether the tactical ladder reflects a tenable developmental sequence in how students may shift from less-to-more advanced tactics, I investigated whether and how students' tactics varied as a function of grade and range.

Table 14

Number of Students per Grade Using Each of Six Tactics

General Tactic Level	0-100			0-1000		
	2	4	6	2	4	6
Advanced						
Proportional Landmarks	1	2	10	0	2	5
Proportional Segmenting	3	3	10	2	6	10
Intermediate						
Landmark Middle	12	13	4	10	10	9
Basic						
Ad Hoc	2	1	0	4	0	0
Ends Only	2	1	0	3	0	0
Guessing	0	0	0	1	2	0

Use. Older students were expected to use advanced tactics more than younger students, because more advanced tactics were hypothesized to emerge in a sequence relative to the level of number-line knowledge or skill needed to select and use those tactics (Figure 4). Older students appeared to use advanced tactics more than younger students (Table 14). The relation between use and number-line knowledge was investigated after characterizing the relations between use of number-line tactics and grade, range, and accuracy.

Relation between tactical use and grade. Table 14 demonstrates that some tactics were used rarely. For example, only three students used a guessing tactic and only on 0-1000. Because some tactics were used less than others, and to test whether use of tactics changed with grade, the six tactics were collapsed into three general tactical skill levels: advanced, intermediate, and basic (Table 14). Separate Fisher's exact tests were conducted for each range to test whether use of the three tactical levels differed among the three grades. The distribution of students across grade differed for the three levels of general tactical skill on each range, *p*s < .01(two-tailed). On the 0-100 line the majority of students in Grades 2 and 4 (60 and 65%, respectively) used intermediate tactics. In contrast, 83% of Grade 6 students used advanced tactics to estimate on 0-100. For the 0-1000 line, half of the Grade 2 and 4 students used an intermediate tactic to estimate. In Grade 2 only two students used an advanced tactic and the rest used basic tactics. In Grade 4 40% of students used advanced tactics, but a few guessed. The majority (63%) of Grade 6 students generally used advanced tactics; the remaining students used intermediate tactics.

Relation between tactical use and range. To assess whether less advanced tactics were used on 0-1000 compared to 0-100, I examined the extent to which the 64 students moved from one tactic to another. Thirty-one students (48%, including 12, 9, and 10 students in Grades 2, 4, and 6, respectively) used the same tactic on both ranges, 11 students (17%, 1, 6, and 4 students in Grades 2, 4, and 6) used a higher tactic on the 0-1000 line, and 22 students (34%, 7, 5, and 10 students in Grades 2, 4, and 6) used a lower tactic on the 0-1000 line. Students who changed tactics were more likely to shift to a lower than a higher tactic for the 0-1000 range (binomial sign test, $p = .02$). Shifting did not vary as a function of grade. The largest shifts were from Landmark Middle to Guessing tactics (two Grade 2 students) and from a Proportional Landmark to an Ad Hoc tactic (a Grade 4 student). The majority of shifts (21 out of 33) were within one tactical level. Finally, controlling for grade, students' tactical use was positively correlated on the two ranges, $r(64) = .49$, $p < .01$.

Summary of tactical use. On the 0-100 line, older students used intermediate and advanced tactics almost exclusively, whereas many younger students used the intermediate tactic, Landmark Middle, and a few younger students used basic tactics such as Guessing or Ends Only (Table 15). For the 0-1000 line a somewhat similar pattern emerged. Grade 6 students used only intermediate and advanced tactics. Grade 4 students tended to use the intermediate tactic, Landmark Middle, or the lower advanced tactic, Proportional Segmenting. In contrast, nearly half the Grade 2 students used the intermediate tactic, Landmark Middle, whereas most other students relied on basic tactics such as Ad Hoc or Ends Only, and only two Grade 2 students used an advanced tactic. These results further highlight the diversity of students' solution operations and illustrate how students' approaches to the task may, with grade, become increasingly sophisticated in a sequence that reflects students' progress toward differentiating and hierarchically integrating number-line knowledge.

Accuracy. The conventional analyses of number-line accuracy reported earlier showed that older students are more accurate on the task than younger students for both ranges. Based on the proposed tactical ladder, using advanced tactics should result in less error whereas using lower tactics should result in higher error. Table 15 shows that, as expected, advanced tactics are associated with less error on the task across both ranges. Even when controlling for grade, use of tactics at a more advanced level was associated with less error for both ranges, $r(61) = -.55$, for 0-100, and $r(61) = -.33$, for 0-1000, ps < .01.

Table 15

Mean PAE of Six Tactics Collapsed Across Grade by Range

Level	Tactic	0-100		0-1000	
		N	Mean % (*SD*)	*N*	Mean % (*SD*)
Advanced					
	Proportional Landmarks	13	4 (2)	7	8 (7)
	Proportional Segmenting	16	4 (2)	18	8 (7)
Intermediate					
	Landmarks and Middle	33	7 (3)	30	9 (5)
Basic					
	Ad Hoc	3	16 (5)	5	21 (4)
	Ends only	3	15 (4)	4	28 (13)
	Guessing	0		3	20 (2)

Note. *SD* = Standard Deviation

Beyond conventional analyses, examining the relation between tactics and error is of interest because it should shed light on how tactics, a measure of procedural knowledge, relate to children's pattern of estimates within a number-line range. For each participant I calculated the discrepancy score between a perfect estimate of the target and the students' actual estimate (Petitto, 1990). Figure 11 shows the mean discrepancy score for each tactic for all students (n = 64) across all grades.

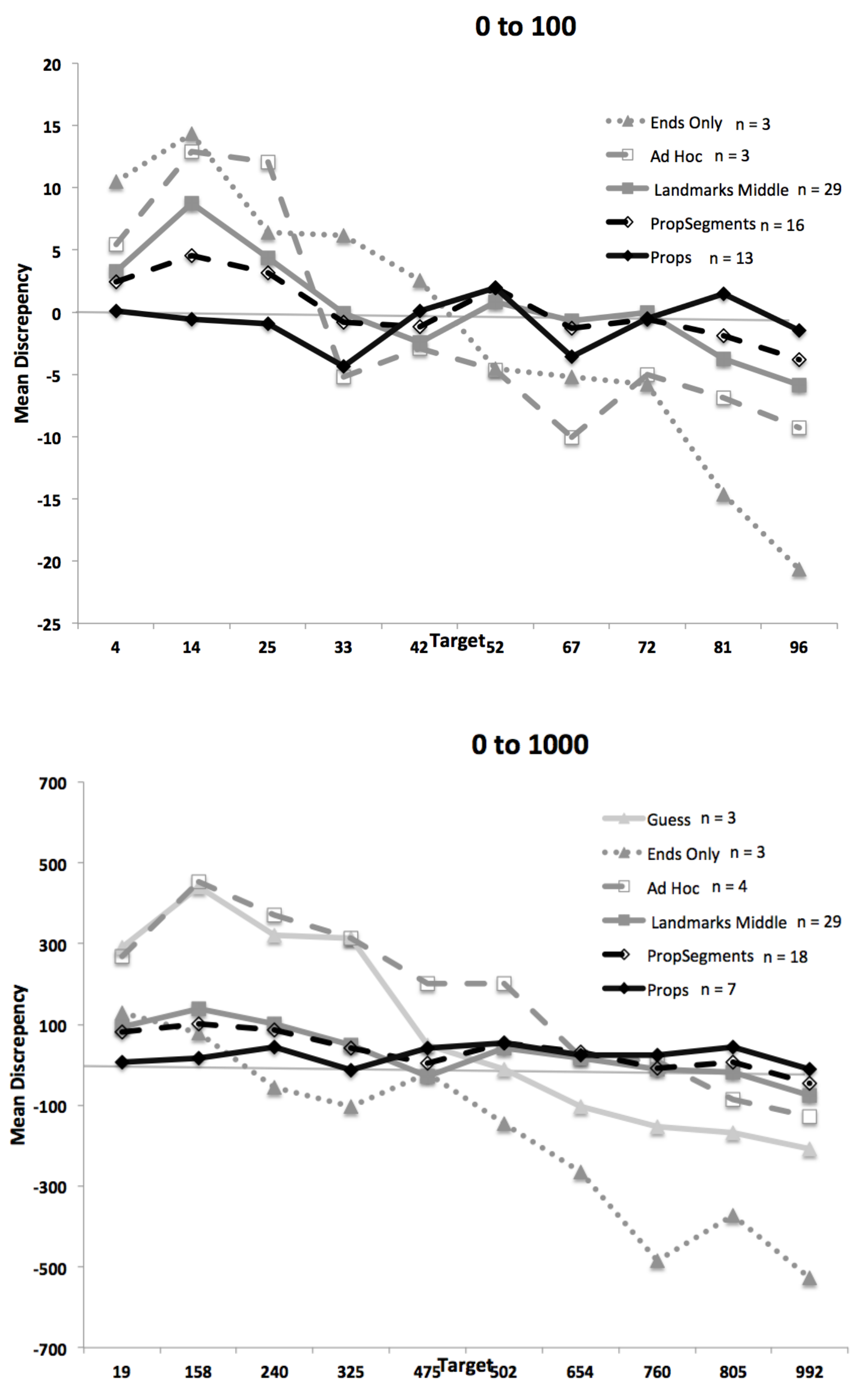


Figure 11. Mean discrepancy for each target by tactic collapsed across grade.

Recall from the conventional analysis that on the 0-100 line students demonstrated a highly linear pattern of estimates (Figure 5). Figure 11, in contrast, points to a more subtle interpretation highlighting the role that procedural knowledge may play in estimation. Despite the appearance of students collectively showing a linear pattern of estimates when averaged together, students approached the line in different ways, resulting in different patterns of discrepancy. More advanced tactics resulted in less discrepancy from the target and more linear estimates. For example, students who used a Proportional Landmark tactic showed little discrepancy compared with students using Ends Only or Ad Hoc tactics. Moreover, students who used a Proportional Landmark tactic were highly accurate near the ends, middle, and ¼ and ¾ landmarks. Students who used the intermediate, Landmark Middle tactic showed little discrepancy near the middle and upper half of the line, and only slight over- and underestimation at the two ends as compared with the advanced tactics.

Conventional analyses on the 0-1000 line showed that older students still had fairly linear patterns of estimates but that Grade 2 students showed a pattern of estimates appearing more curvilinear (Figure 5). As with the 0-100 line, Figure 11 illustrates the need for a more nuanced interpretation of what underlies number-line estimation that includes consideration of students' procedural knowledge. As on the 0-100 line, students who used a Proportional Landmarks tactic on the 0-1000 line showed little discrepancy from a perfectly linear pattern compared with students who used an Ends Only tactic. On the 0-1000 line the differences in over- and under-estimation stand out more for particular tactics. Students classified as Guessing on the 0-1000 line showed overestimation on the lower half of the line but were accurate around the middle and underestimated on the upper portion of the line. In contrast, students who used an Ad Hoc tactic looked similar to students using a Guessing tactic on the lower half of the line but showed the most discrepancy at the middle and less underestimation on the upper half of the line. Finally, students who used Ends only overestimated on the lower half of the line but looked fairly similar to some of the more advanced tactical users with little discrepancy around the middle and little underestimation on the upper half of the line.

Only one tactic, Landmark Middle, was used frequently enough across grades and in both ranges to examine whether, for the same tactic, accuracy for each range changed as a function of grade. Figure 12 illustrates the grade-related differences in discrepancy of the Landmark Middle tactic. To investigate the effect of grade, separate one-way ANOVAs for each range were run on mean PAE for students using the Landmark Middle tactic. On the 0-100 line, older students showed less error than younger students (*M*s = 7%, 7%, and 4% for Grades 2, 4, and 6, respectively), $F(2,26) = 3.14$ $p = .06$, $\eta_p^2 = .20$. The four Grade 6 students using a Landmark Middle tactic showed almost no discrepancy in their answers at the two ends, and estimated accurately near the middle. Students in Grades 2 and 4 had nearly overlapping patterns of overestimation on the lower half of the 0-100 line and on the upper half of the line, Grade 2 students underestimated more than Grade 4 students.

On the 0-1000 line older students executed the Landmark Middle tactic better than younger students as shown by the decrease in error with grade (*M*s = 12%, 8%, and 5% for Grades 2,4, and 6, respectively), $F(2, 26) = 4.61$, $p < .05$, $\eta_p^2 = .26$. Students in all grades estimated with high accuracy around the midpoint but Grade 2 students overestimated the most on the lower half, and also underestimated the most on the upper half. Grade 6 students overestimated on both halves and underestimated slightly near the end of the line. The pattern of Grade 4 students was similar in shape to that of the Grade 6 students but with slightly more overestimation at the low end and more underestimation at the high end. The implication of these results is that although students used similar kinds of procedural knowledge there are still differences in execution across grades. These grade differences could be driven by differences in the execution of the Landmark Middle tactic such as the accuracy of anchoring to the middle in the first place, and the ability to accurately adjust from an anchor.

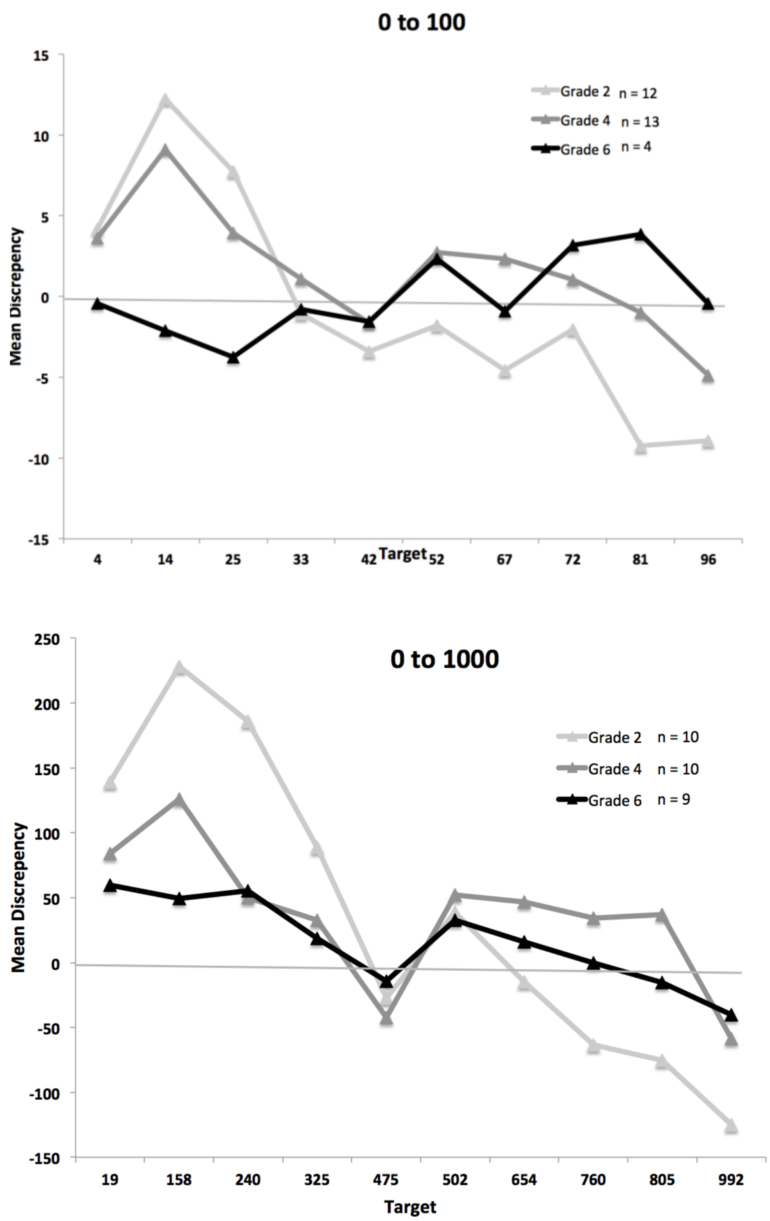


Figure 12. Mean discrepancies of Landmark Middle tactic by range and grade.

Collectively, the results from examining discrepancy across targets as a function of tactic illustrate that even when all students appear to have a linear pattern of estimation, not all students are demonstrating the same level of procedural knowledge when estimating. Second, close inspection of the Landmark Middle tactic reveals clear grade-related differences. In this study, equations were not fit to patterns of estimates because only 10 trials per range were collected with self-reports, compared with studies where a model-fitting approach is taken where models are fit to 20 or more trials per range. The clear variability, however, in children's number-line tactics leads me to suggest that current model-fitting approaches fail to account for a number of features of children's number-line estimation.

First, models of number-line estimation that posit that students use reference points do so with the assumption that students use reference points accurately. In some cases, as with the endpoints, such an assumption may be warranted, but in the case of a midpoint, for example, students may not use the middle accurately. For example, in this study, I clearly observed eight students, four in Grade 2, and two in each of Grades 4 and 6, who indicated that the physical middle of the 0-1000 line was either 100 or 50 (the former was more common). This observation warrants that a full account of the development of number-line estimation using a model-fitting approach must include some discussion of how students come to use accurate, mathematically correct reference points in their number-line estimations.

Second, I suggest that models of number-line estimation must account for students' accuracy in subsequent processes after anchoring such as iteratively anchoring and adjusting. For example, to estimate 72 on a 0-100 line, a student might accurately anchor to the middle followed by counting by tens to 70 (an example of iteratively anchoring), and then count by ones from 70 to find 72, but may use tens and ones increments that are too large for the scale of the line. As a result, the source of their overestimation is clearly related to the accuracy of the processes that follow their initial anchoring. Again, based on the task analysis and the coding of adjustment processes, an account of how children's number-line estimation develops must also include a way to accurately capture the variability in students' anchors and adjustments. Arguably, the idea of some predictor in a model to account for variability in potential iterative anchoring processes, and adjusting could be simply stated as, for example, an "adjustment predictor". I would caution against such an oversimplification, however, because to include such a predictor would be to assume that everyone adjusts equally from every anchor or reference point on the line.

Inspection of the bottom panel of Figure 12, however, clearly demonstrates that the amount of over- and under-estimation is not always equal for different parts of the line, nor always in a direction indicative of a cyclic model such as the two- and three-anchor models. That is, all of the students using a Landmark Middle tactic showed a pattern of overestimation on the lower half of the line (as expected). On the upper half of the line however, only Grade 2 students showed something close to the expected pattern of underestimation and, critically, the amount of underestimation is less than the amount of overestimation: There is an apparent asymmetry to the pattern of over- and under-estimation. Because all of these students are known to be using the same tactic, that is using the ends and middle, future models of number-line estimation will have to account for the other processes such as adjustment that may lead to such variability in executing these tactics.

Relation of tactics to number-line knowledge. LeFevre et al. (2013) demonstrated a link between accuracy on number-line estimation and conceptual knowledge of numbers, and both Ashcraft and Moore (2012) and White and Szűcs (2012) hypothesized that knowledge about the number line may be related to approaching the line using more advanced tactics. What is unique in this study is that each student has an index of her explicit conceptual knowledge about the number line, K_{num}, as well as an index of her tactical level within a range for both 0-100 and 0-1000. If conceptual and procedural knowledge develop iteratively in number-line estimation, having a high index of conceptual knowledge should be positively correlated with using an advanced tactic. For each tactical level on each range, mean K_{num} scores are shown in Table 16. Students' mean K_{num} was higher for more advanced tactical levels in both ranges. Controlling for grade, higher K_{num} was associated with use of higher levels of tactics for 0-100, $r(62) = .45$, $p < .01$, but not for 0-1000, $r(62) = .16$.

There are at least two potential interpretations for why the expected relation may have been found only for the 0-100 and not for the 0-1000 line. First, the K_{num} measure may have been insensitive to measuring what students know about the 0-1000 line because the items on K_{num} focused more on the 0-100 than the 0-1000 line. Only two questions explicitly tapped into knowledge about the 0-1000 line. Second, students' knowledge of number lines may be highly compartmentalized (Hiebert & Lefevre, 1986) such that certain knowledge is tied to a particular number-line scale. For example, students may know the location of proportions in the context of a particular number-line (such as 0-100) but be unable to apply this knowledge about general properties of the number-line to another scale. The measure of explicit number-line knowledge used here did not include questions that would shed light on the second possibility. Despite the absence of a statistical effect the expected pattern of more advanced tactics being associated with a higher K_{num} score was found for both ranges (Table 16).

Table 16

Mean K_{num} of Six Tactics Collapsed Across Grade by Range

Tactic	Range			
	0-100		0-1000	
	N	Mean K_{num}	*N*	Mean K_{num}
Proportional Landmarks	13	7.48	7	6.68
Proportional Segments	16	6.35	18	6.00
Landmark Middle	29	4.69	29	5.76
Ad Hoc	3	3.33	4	3.25
Ends Only	3	3.33	3	3.33
Guess	0	-	3	3.33

Measures of Number-line Knowledge and Math Achievement

As described in the introduction, several studies demonstrate that the more linear and accurate a student's estimates on number lines, the higher his scores on tests of general mathematics achievement (Ashcraft & Moore, 2012; Booth & Siegler, 2006; Sasanguie et al., 2013; Siegler & Booth, 2004; Träff, 2013), or the higher his grade in math (Schneider, Grabner, & Paetsch, 2009). The relations among additional measures of number-line knowledge tested in the present study, procedural and explicit conceptual knowledge, and math achievement were tested. Two measures of mathematics achievement from the WJ-III, Math Fluency and Applied Problems, were correlated with four measures of number-line knowledge: two measures of implicit, conceptual knowledge used in previous studies, R^2_{Lin}, and mean proportion of absolute error on 0-100 and 0-1000 lines; an explicit conceptual number-line measure indexed by K_{num}; and a procedural measure indexed by tactics on the 0-100 and 0-1000 lines (Table 17). For the mathematics achievement tests students' raw scores were used because the aim was to examine relations based on children's overall mathematical knowledge rather than their mathematical knowledge relative to children of the same age.

Table 17

Correlations Among Number-line Measures and Mathematics Achievement

Measures	1	2	3	4	5	6	7	8	9
1. Linear R^2 of 100	__	.500**	-.894**	-.419**	.598**	.582**	.484**	.438**	.514**
2. Linear R^2 of 1000	.349**	__	-.514**	-.859**	.480**	.632**	.569**	.507**	.545**
3. Mean PAE 100	-.872**	-.277*	__	.485**	-.676**	-.618**	-.630**	-.612**	-.667**
4. Mean PAE 1000	-.250	-.802**	.252*	__	-.453**	-.522**	-.668**	-.547**	-.575**
5. Tactic on 100	.420**	.226	-.426**	-.206	__	.494**	.620**	.508**	.521**
6. Tactic on 1000	.477**	.492**	-.476**	-.351**	.290*	__	.433**	.428**	.468**
7. K_{num}	.217	.303*	-.284*	-.506**	.267*	.121	__	.745**	.757**
8. Math Fluency	.127	.147	-.229	-.264*	.010	.069	.391**	__	.704**
9. Applied Problems	.257*	.265*	-.342**	-.352**	.044	.187	.443**	.284*	__

Note. Simple correlations (n = 64) are shown above the diagonal and partial correlations, controlling for BIA and grade (df = 60), are shown below the diagonal. *p <.05; ** p<.01.

As found in previous literature, the measures of implicit conceptual number-line knowledge were correlated with both math achievement measures. Measures of procedural and explicit conceptual number-line knowledge were also correlated with both measures of math achievement such that the more advanced the student’s tactic, and the greater the student’s explicit number-line knowledge, the higher the student’s scores were on the math achievement tests. Measures of number-line knowledge were correlated with one another. All measures were correlated with grade and a standardized measure of intelligence, the BIA.

After controlling for BIA and grade, measures of implicit conceptual number-line knowledge remained correlated with Applied Problems, consistent with the expected pattern (Ashcraft & Moore, 2012; Sasanguie et al., 2013; Siegler & Booth, 2004). Only mean PAE on the 0-1000 line remained correlated with Math Fluency, a result similar to a finding by Träff

(2013). Controlling for BIA and grade, correlations among the measures of number-line knowledge were attenuated and revealed a pattern of relations consistent with the idea that the development of procedural and conceptual knowledge are closely linked and may emerge iteratively and gradually (Rittle-Johnson et al., 2001). For example, the ability to use half or proportions as reference points on the number line may be linked to an explicit understanding of understanding specific proportions.

Relations among measures of number-line knowledge. Other researchers have suggested that patterns of estimates on the number-line task should reflect the tactics students used to solve number-line estimation problems (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Rouder & Geary, 2014; Siegler & Opfer, 2003; White & Szűcs, 2012). Tactics were expected to correlate with the two measures of implicit conceptual number-line knowledge. Table 17 illustrates the expected relation: The more linear and accurate a student's estimates on the 0-100 line, the higher a student's tactical level on both the 0-100 and 0-1000 lines. Moreover, the more linear and accurate a student's estimates on the 0-1000 line, the higher the student's tactical level on the 0-1000 line. Because tactics index students' procedural number-line knowledge trial-by-trial across each number-line range, these results represent a categorical link between greater linearity and the use of advanced tactics to estimate across a number line.

Next, the relation between implicit and explicit conceptual number-line knowledge was tested. The two measures of conceptual knowledge were correlated, but not always strongly, suggesting that the two measures may tap different kinds of conceptual knowledge. The more accurate a student's estimates were on both lines, the higher the student's K_{num} score. The more linear a student's estimates were on the 0-1000, but not the 0-100 line, the higher the student's K_{num} score. Finally, the higher a student's tactical level on the 0-100, but not on the 0-1000 line, the higher the student's K_{num} score. This result links procedural number-line knowledge to explicit number-line knowledge but it is surprising that the relation did not hold for both ranges. As noted earlier, one explanation may be that the explicit measure of number-line knowledge focused almost exclusively on the 0-100 line with just a few questions about scale changes involving 0-1000. Had more questions about the 0-1000 line been included in K_{num}, such as the locations of proportions on 0-1000, tactics on 0-1000 may have correlated with a more comprehensive K_{num} measure.

Relation of number-line knowledge to math achievement. A recent longitudinal study demonstrated that the link between number-line estimation and math achievement may be indirect and better characterized as a link between spatial reasoning and general number knowledge (LeFevre et al., 2013). Schneider and colleagues (2009) also found that conceptual number knowledge about decimal fractions was associated with having a higher math mark for students in Grades 5 and 6, and Ashcraft and Moore (2012) found that students classified as linear responders scored higher on a measure of mathematics achievement than students classified as exponential responders (that is, having a more curvilinear pattern of number-line estimates).

Controlling for grade and BIA, tactics on both 0-100 and 0-1000 were not correlated with either measure of math achievement. Not finding a relation between number-line tactics and math achievement may be the result of number-line tactics being highly specific to the number-line domain and students recruiting specific procedural knowledge not tapped by tests of exact calculations (Math Fluency) or general applied problem solving (Applied Problems). The greater a student's explicit conceptual number-line knowledge, the higher his or her scores were on both math achievement tests.

To test whether K_{num} contributes uniquely to performance on Applied Problems beyond both implicit conceptual measures of number-line knowledge, separate regressions were run for each range. For both ranges, K_{num} was uniquely associated with performance on Applied Problems after controlling for both linear fit and mean PAE, $\Delta R^2 = .16$, $F(1,60) = 27.14$ for the 0-100 line, and $\Delta R^2 = .25$, $F(1,60) = 37.23$ for 0-1000 line, $ps < .01$. Even when including grade and BIA as predictors in addition to linear fit and mean PAE for each range, K_{num} continued to uniquely account for additional variation in scores on Applied Problems. This result indicates that potentially the kinds of number-line knowledge that generalize to other kinds of numerical knowledge are not captured exclusively by measuring accuracy or linearity on number-line estimation, but that explicit conceptual knowledge about the number line reflects, over and above implicit conceptual knowledge, important units of numerical knowledge related to overall achievement in mathematics.

Exploration of Tactical Adjustments

One motivation for studying children's thinking on the number-line task is that it provides an opportunity to investigate how children adjust or adapt their thinking (LeFevre et al., 2013; Newman & Berger, 1984; Siegler & Booth, 2005). Because there were no expectations about how students might adjust their tactics across ranges, a data-driven approach was adopted to identify patterns of tactical adjustments. Seventeen patterns of tactical adjustments were observed as students shifted from the 0-100 to the 0-1000 line (Table 18). Of 64 students, 48% maintained their tactic across the two ranges, compared with 34% who shifted down to less advanced tactics. The remaining 17% of students shifted up to more advanced tactics. Of the 17 tactical adjustments, Grade 6 students used eight different adjustments and students in Grades 2 and 4 each used 10 adjustments. The specific adjustments within each grade are documented in Table 18.

Table 18

Seventeen Tactical Adjustments by Grade

Tactic Shift			Grade			
From	To	Shift	2	4	6	Total
Ends	Guess	Down	0	1	0	1
Ends	Ends	Maintain	2	0	0	2
Ad Hoc	Ad Hoc	Maintain	1	0	0	1
Ad Hoc	Proportional Segments	Up	1	1	0	2
Landmark Middle	Guess	Down	1	1	0	2
Landmark Middle	Ends	Down	1	0	0	1
Landmark Middle	Ad Hoc	Down	2	0	0	2
Landmark Middle	Landmark Middle	Maintain	8	7	1	16
Landmark Middle	Proportional Segments	Up	0	4	3	7
Landmark Middle	Proportional Landmarks	Up	0	1	0	1
Proportional Segments	Landmark Middle	Down	2	2	4	8
Proportional Segments	Proportional Segments	Maintain	1	1	5	7
Proportional Segments	Proportional Landmarks	Up	0	0	1	1
Proportional Landmarks	Ad Hoc	Down	1	0	0	1
Proportional Landmarks	Landmark Middle	Down	0	1	4	5
Proportional Landmarks	Proportional Segments	Down	0	0	2	2
Proportional Landmarks	Proportional Landmarks	Maintain	0	1	4	5

To index the adaptivity of tactical adjustments, each student's change in error from 0-100 to 0-1000 was calculated (Lemaire et al., 2000). If change in a student's error is minimal, presumably she adjusted her tactic and thinking in an adaptive way as task demands increased. In contrast, a large increase in error is presumed to reflect a maladaptive or ineffective adjustment, or failure to adjust when an adjustment might have been helpful.

After excluding one outlier (mean Δ PAE = 43%), students were grouped into one of three levels of change in error as shown in Figure 13. The group showing *minimal* change consisted of 32 students with change scores (PAE for 0-1000 minus PAE for 0-100) less than or equal to the median, 2%. The group showing *moderate* change included the 15 students with change scores between the median and one standard deviation above the median, 7%. The remaining 16 students showed *substantial* change (>7%).

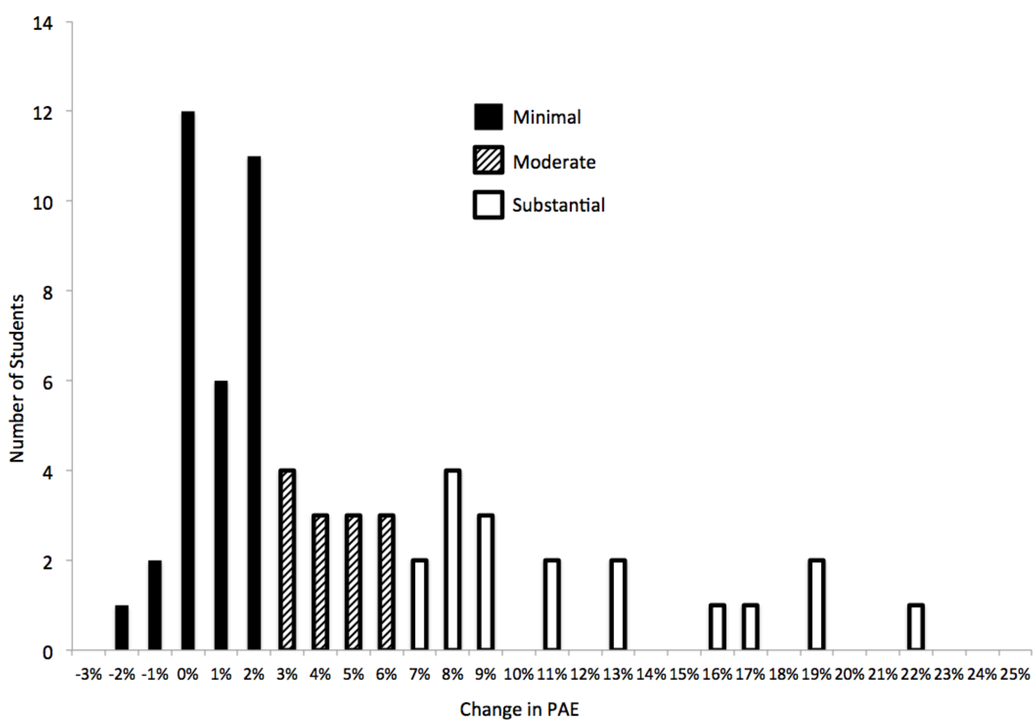


Figure 13. Distribution of students across three levels of mean Δ PAE from 0-100 to 0-1000 line.

Table 19 shows that students with minimal or moderate change in error who initially used tactics involving the midpoint on the 0-100 line *never* shifted down to a tactic that did not include the midpoint. In contrast, of the 14 students initially using a tactic that included the midpoint on the 0-100 line and who showed substantial change in error across the number lines, more than a third (five students) shifted down to a tactic that did not include use of the midpoint. To better understand which tactical adjustments are more adaptive, shifts were examined in relation to changes in error (Table 20). Five preliminary inferences were made based on comparisons among cells with adequate numbers (at least four students per cell).

Table 19

Number and Nature of Students' Adjustments as a Function of Error Level

Tactic			Change-in-Error		
From 0-100	To 0-1000	Shift	Minimal	Moderate	Substantial
Proportional Landmarks					
	Ad Hoc	Down	0	0	1
	Landmark Middle	Down	4	1	0
	Proportional Segments	Down	2	0	0
	Proportional Landmarks	Maintain	4	0	1
Proportional Segments					
	Landmark Middle	Down	4	4	0
	Proportional Segments	Maintain	5	2	0
	Proportional Landmarks	Up	0	0	1
Landmark Middle					
	Guess	Down	0	0	2
	Ends	Down	0	0	0
	Ad Hoc	Down	0	0	2
	Landmark Middle	Maintain	7	5	4
	Proportional Segments	Up	4	0	3
	Proportional Landmarks	Up	1	0	0

Table 20

Nature of Students' Directional Change in Tactics by Change in Error

Tactic on 0-100	Amount of Error	Directional Change in Tactic		
		Maintain	Up	Down
Proportional Landmarks				
	Minimal	4		6
	Moderate	0		1
	Substantial	1		1
Proportional Segments				
	Minimal	5	0	4
	Moderate	2	0	4
	Substantial	0	1	0
Landmark Middle				
	Minimal	7	5	0
	Moderate	5	0	0
	Substantial	4	3	4

First, the four students who adapted (showed a minimal change in error) by maintaining the Proportional Landmarks tactic were in the same grade on average as the six students who adapted by downshifting from the Proportional Landmarks tactic (*M*s = 5.5 and 5.67, respectively). What distinguished the two groups was their amount of explicit number-line knowledge: students who downshifted had a higher level of explicit number-line knowledge (K_{num} *M* = 8.48) than students who maintained the tactic (K_{num} *M* = 5.67).

Second, of the 15 students who used a Proportional Segment tactic on the 0-100 line and showed either minimal (nine students) or moderate (six students) increases in error, none upshifted to the Proportional Landmark tactic. The one student who upshifted from the Proportional Segments to the Proportional Landmark tactic showed a substantial increase in error on the 0-1000 line. Among the nine students who showed a minimal change in error from the 0-100 to the 0-1000 line, students differed in terms of grade and explicit number-line knowledge depending on whether they maintained the tactic or downshifted. The five students who adaptively maintained the Proportional Segments tactic were at a higher grade level (*M* = 5.2) and had a slightly higher level of explicit number-line knowledge (K_{num} *M* = 7.2) than the four students who adaptively downshifted (Grade *M* = 4.5 and K_{num} *M* = 6.75).

Third, of the eight students who started with the Proportional Segments tactic and downshifted, four demonstrated a minimal increase in error and four showed a moderate increase in error. Grade level did not differ between these two groups (*M* = 4.5 for both) but the four adaptive downshifters had a slightly higher level of number-line knowledge than the other four students (K_{num} *M* = 6.75 and 6.12, respectively).

Fourth, the 12 adaptive students who initially started from the Landmark Middle tactic either maintained the tactic (seven students) or upshifted (five students); they never downshifted. The seven adaptive maintainers of the Landmark Middle tactic were in a lower grade on average than the five students who adaptively upshifted (*M*s = 3.43 and 4.8, respectively). Both groups of adaptive students showed similar levels of explicit number-line knowledge (K_{num} *M* = 5.3 for both).

Finally three groups of students maintained the Landmark Middle tactic with varying degrees of success: Seven students showed a minimal increase in error, five showed a moderate increase, and four showed a substantial increase. The seven adaptive maintainers of the Landmark Middle tactic were in a slightly higher grade level on average (*M* = 3.43) and had higher levels of explicit number-line knowledge (K_{num} *M* = 5.26) than the five students who showed a moderate increase in error (Grade *M* = 3.20 and K_{num} *M* = 3.6). The four students who demonstrated a substantial increase in error when maintaining the Landmark Middle tactic had the lowest grade level on average (*M* = 2.50), and showed a level of explicit number-line knowledge (K_{num} *M* = 5.25) similar to that of the adaptive maintainers.

From this collection of results it appears that the term adaptive cannot be limited to a single tactical adjustment. Instead, a variety of tactical adjustments can lead to adaptive performance on the number-line task. Whether a student's tactical adjustment is adaptive, and the conditions under which students adjust their tactics, seems to depend on a number of factors unique to the individual. Such factors may include students' initial tactic on the 0-100 line, grade, and level of explicit number-line knowledge. The number of students per cell in this exploration is low but some tentative implications from these results can be highlighted.

First, it does not appear that any one specific tactical adjustment will guarantee a student will demonstrate minimal change in error. For example, maintaining the Landmark Middle tactic across the two ranges resulted in a range of outcomes from minimal to substantial increases in error. In the case of the eight students who downshifted from a Proportional Segments tactic, half demonstrated minimal increase in error and the rest a moderate increase. Therefore not only does the type of tactical adjustment matter but almost certainly the execution of procedures within a tactic also has some bearing on the effectiveness of the tactical adjustment. Moreover, a particular kind of movement on the tactical ladder (maintain, up, or down) is not uniquely associated with minimal change in error nor is starting from one particular tactic. What is most striking about these observations is the diversity of potentially adaptive moves students may make.

Another implication of these results is that adaptivity may be constrained by a number of characteristics unique to the individual. For example, students who used a Proportional Landmarks tactic on the 0-100 line and adapted (minimal increase in error), only ever maintained their tactic or downshifted. Whether these students maintained their tactic or shifted down appeared to be related to their level of explicit number-line knowledge; not grade level. The reverse pattern was found for students who used the Landmark Middle tactic on the 0-100 line and adapted (minimal increase in error). Whether those students using Landmark Middle maintained or upshifted appeared to be related to grade; not their level of explicit number-line knowledge. These results illustrate that multiple factors may constrain adaptivity differently as a function of age, K_{num} , and starting tactic.

As a result it might be useful to consider adaptivity as a multi-dimensional space in which people can achieve adaptivity via different routes. Research programs about adaptivity might include not only questions of whether a student is adaptive or adaptive to a certain degree, but also questions designed to identify and measure a range of factors that may be associated with cognitive profiles or constellations of characteristics that are adaptive.

General Discussion

How children come to understand number-line estimation has been a topic of much research in the past decade. Most researchers have focused on characterizing and explaining number-line estimation by fitting equations to children's estimation patterns (e.g., Ashcraft & Moore, 2012; Barth & Paladino, 2011; Booth & Siegler, 2006; Bouwmeester & Verkoeijen, 2012; Opfer & Siegler, 2007; Rouder & Geary, 2014; Siegler & Opfer, 2003; Siegler & Booth, 2004; Slusser et al., 2013; Young & Opfer, 2011). Initially, number-line estimation was conceptualized in terms of children beginning with a compressed representation of the number line that, with development and experience, shifted to a linear representation (Siegler & Opfer, 2003; Siegler & Booth, 2004). Mounting evidence, however, suggests that the shift from a compressed-to-linear representation fails to fully account for (a) how children approach the task or the kinds of procedures used to estimate (Barth & Paladino, 2011; Slusser et al., 2013), (b) variability as a function of target position on the number line (Ashcraft & Moore, 2012), and (c) individual variability in estimation patterns (Bouwmeester & Verkoeijen, 2012; Rouder & Geary, 2014).

A recent, sophisticated, quantitative approach to fitting equations to children's patterns of estimates on the 0-100 line is Rouder and Geary's (2014) longitudinal study of number-line estimation with over 200 students. Rouder and Geary's work is notable for several reasons. First, they assessed each student's patterns of estimates on a 0-100 number line once in each of Grades 1 through 5. Consistent with cross-sectional studies of number-line estimation Rouder and Geary found that students' estimates became more accurate over five years. Rouder and Geary analyzed students' patterns of estimates by fitting equations that included model-specific log-odds transformations for specific targets on the line. By doing this, Rouder and Geary assert, they were able to "linearize the means and stabilize the variances" (p. 5) and therefore isolate and analyze the between-subject variability in children's estimation patterns. Using this approach Rouder and Geary's results echoed earlier findings (Barth & Paladino, 2011; Slusser et al., 2012) that changes in patterns of estimates may be understood in terms of a progression toward proportional reasoning. Rouder and Geary found evidence for this progression by fitting one-, two- and three-anchor models to individuals' data across five test times and selecting the best fitting model for each individual-by-grade combination.

Rouder and Geary (2014) argued that children initially use one anchor, 0, to estimate, which corresponds psychologically to a compressed model of the number line and mathematically to a one-anchor model. As they develop, children begin to use both ends of the line to estimate because they are beginning to approach the number line as a bounded task, as proposed by Barth and colleagues (Barth & Paladino, 2011; Slusser et al., 2012). Use of the ends corresponds, mathematically, to a two-anchor model. Finally, by drawing upon their ability to think proportionally, children use three anchors to estimate: the ends *and* middle. This corresponds, mathematically, to the three-anchor model. Rouder and Geary (2014) suggested that the more anchors a student uses, the more constrained, and therefore more accurate, students are in estimating targets.

Supporting the idea of increasing linearity in number-line estimation as being a progression from the use of fewer-to-more anchors, Rouder and Geary's (2014) results showed that the equation reflecting a one-anchor model fit best the accuracy patterns of 20% of students in Grade 1 and less than 9% of the students by Grade 5. They also found that nearly two-thirds of the students' estimates in Grade 1 were best fit by an equation reflecting a two-anchor model and that by Grade 2, a plurality of students had a pattern of estimates best fit by an equation reflecting a three-anchor model. By Grade 5, more than half the children showed a pattern of estimates best fit by an equation reflecting a three-anchor model. Rouder and Geary's results paint a picture of a transition, for the 0-100 line, occurring between Grade 1 and 2 in which students move from using two anchors, the ends, to using three anchors, the ends and midpoint. Though Rouder and Geary speculated that students might come to use additional anchors such as a quarter and three quarters, they found no evidence for this using their model-fitting approach.

Measuring and Mapping Number-line Knowledge

The methodological approach taken in this dissertation complements previous research and yields results that are consistent with, and extend, recent findings in number-line estimation (Ashcraft & Moore, 2012; Rouder & Geary, 2014; White & Szűcs, 2012). Two general questions were posed: "What does the child do when estimating?" and "What does the child know about the number-line?" These two questions were addressed by (a) expanding the focus of assessment from accuracy to include students' procedural knowledge and explicit conceptual knowledge of number-line estimation and (b) carefully considering variability as a function of individuals as well as targets within a range.

Self-reports were used to verify the suspected use of anchoring processes in number-line estimation (Ashcraft & Moore, 2012; Rouder & Geary, 2014; Schneider et al., 2008; White & Szűcs, 2012) and to investigate whether students use other processes to estimate. Complementing the methodological approaches of other researchers who have worked to capture variability in estimation as a function of target (e.g., Ashcraft & Moore, 2012; Schneider et al., 2008; White & Szűcs, 2012), a task analysis of number-line estimation was proposed and used to organize and code students' self-reports of estimation. As a result and consistent with previous research (e.g., Ashcraft & Moore, 2012; Schneider et al., 2008; White & Szűcs, 2012), students' use of a variety of anchors was confirmed. As suggested by other researchers (e.g., Ashcraft & Moore, 2012; Barth & Paladino, 2011; Schneider et al., 2008; Siegler & Opfer, 2003; White & Szűcs, 2012) the most frequently used procedures involved students anchoring to the ends and middle.

Adding to the existing number-line literature is the discovery of substantial variability in the other kinds of anchors students used. Although anchoring to the ends or middle made up half of the procedures students used, other anchoring patterns were identified. As suggested by some researchers (Rouder & Geary, 2014), not only did students use ends and middle as anchors but, in some cases, these anchors were used together, to help students anchor to a segment of the line. For example, students anchored to the upper or lower half of the line, and some students located additional anchors within a segment, such as the middle of the upper or lower half. Students also reported using other kinds of proportions such as quarter and three-quarters and, in some instances, a third, or constructed proportional anchors such as fifths or tenths. Finally, students generated and constructed anchors not previously mentioned in the literature by anchoring to whole decades (“the 20s”) or referencing an anchor from a previous trial (“I thought 14 was here before so here is 14”).

Students also used a variety of adjustment processes, after anchoring, to refine their estimations. Rouder and Geary (2014) suggested that students might calculate after anchoring. Analyzing processes revealed that some students did make a quantitative adjustment, similar to calculation (“go up 2” or “add 2”). Students, however, had other kinds of adjustments at their disposal. For example, students used qualitative adjustments (“up a bit”), guessed, constructed a proportion between two anchors, or recalled a previous adjustment (“last time 4 was about here from 0 so 96 is here from 100”). Reformulation processes such as changing the scale of the line or target were also observed, although infrequently. This level of detail in coding the self-reports reveals that students can recruit and assemble multiple processes into a variety of procedures to make a single estimate.

Another novel finding in this dissertation is the variety of solution procedures students used to estimate. Out of 40 anchor-by-adjustment combinations, plus guessing, 38 different procedures were observed. Focusing on those procedures used by more than a single student on one or two trials, 24 different procedures accounted for how students estimated on 96% of targets. Documenting the procedures students used for each target enabled the classification of each individual’s number-line estimation tactic for each number-line range. Until now, researchers have generally speculated about the nature of the cognitive processes that underlie number-line estimation (but see Schneider et al., 2008; Sullivan et al., 2011) based on results obtained by fitting equations to patterns of accuracy data. In the present study tactics reflect the procedures students’ used across targets within a range, and therefore additional evidence and new insights were generated about the cognitive processes students used to estimate.

By assessing the procedures students used as a function of target quintile, the results of this dissertation are consistent with and confirm the use of various tactics on the number line. First, though not called a tactic in the literature, the three-anchor model, or anchoring to the ends and middle, maps onto the Landmark Middle tactic. The Landmark Middle was found to be the most commonly used tactic for both ranges. The Landmark Middle tactic was, however, only one of six identifiable tactics typically developing children in Grades 2, 4, and 6 used to estimate. Also identified using the framework developed here was the Ends tactic, which maps onto what other researchers have called the two-anchor model (e.g., Rouder & Geary, 2014). Consistent with the assertion that younger students use fewer anchors to estimate compared with older students, it was found in the present study that the Ends-only tactic was used almost exclusively by Grade 2 students. Other tactics identified included the Proportional Landmark tactic (analogous to what Rouder and Geary might call a five-anchor model) and the Ad Hoc tactic (comprising other anchors constructed by students).

Second, the identification of tactics leads to expansion of the idea that students progress from using fewer to more anchors as they learn to reason proportionally. The tactical ladder, as a model of what changes over time in number-line estimation, is organized to depict development as a progression toward differentiating parts of the line and integrating across the line. Compared to previous models of change in number-line estimation, the tactical ladder includes more specificity, accounts for potential extremes (guessing and using pure proportional thinking) and also includes the class of tactics (Ad Hoc) composed of students' own generated and constructed anchoring patterns. Thus, extending the results of Rouder and Geary and the number-line literature more generally is the discovery that at least six tactics can be ordered, speculatively, into a developmentally tractable sequence.

Third, identifying what students were doing across a number-line range allowed for the investigation of whether accuracy for particular tactics varied with grade and range. Only one tactic, Landmark Middle, was used frequently enough across all three grades and in both ranges to examine whether there are differences in accuracy as a function of range and grade when students are using the same tactic. The results demonstrated that even when students rely on the same tactic, older students execute the tactic better than younger students. One explanation of what might change with grade is the ability to use the middle more accurately. Moreover, there may be differences in the ability to invoke adjustment processes that change as a function of age. By mapping number-line knowledge in detail, it was possible to identify and investigate more precisely the processes of estimation that may develop, such as anchoring and adjusting.

Fourth, in addition to characterizing the processes, procedures, and tactics of number-line estimation, children's explicit conceptual number-line knowledge was found to increase with grade and was positively related to the ability to use more advanced tactics to estimate on both ranges. That is, students with more knowledge about explicit number-line concepts were able to construct more conceptually advanced problem-solving tactics. The implication of this finding is not that one kind of knowledge is necessarily more central to development than another, but that understanding what develops and how in number-line estimation entails describing relations in a robust and complex web of knowledge.

Finally, Rouder and Geary's longitudinal data demonstrated that students' generally became more accurate and linear in their estimates over time. Rouder and Geary suggested that, with age, students adjusted their approach to estimating by first using just one, then two, followed by three anchors to estimate. In the present study, the developmental progression from fewer to more anchors was verified and further refined. As found on other mathematical tasks (e.g., Lemaire & Siegler, 1995; Luwel, Siegler, & Verschaffel, 2008; Siegler & Stern, 1998) and in other domains (e.g., Rittle-Johnson & Siegler, 1999), results from the present cross-sectional study illustrate that students adjusted their solution procedures trial-by-trial. Moreover, the number-line task and multi-method approach to measurement gave rise to the opportunity to look at the extent to which students adjusted their use of solution procedures trial-by-trial within a range, and across ranges.

Number-line Estimation as a Vehicle for Studying Adaptivity

In their study of multiplication, Lemaire and Siegler (1995) characterized the strategic aspects of children's cognition in terms of using new strategies, using the most efficient strategies more often, executing strategies more efficiently, and choosing strategies adaptively. For Lemaire and colleagues, adaptivity is indexed by the correlation between the child's use of a strategy and the problem characteristics (Lemaire, et al., 2000; Lemaire & Lecacheur, 2002). In the present number-line estimation study adaptive thinking was characterized by taking a slightly different tack because the number-line task itself affords inventiveness across trials within a range. As a result, the current approach to adaptivity was to analyze students' solution operations at multiple levels, from processes through procedures to tactics, and to create a kind of "zoom" feature in the analyses to understand how students were thinking flexibly at different levels. This approach, in the context of the number-line estimation task, has the potential to be the kind of single complex task, accessible to even young children, around which experimental and theoretical studies can be devised to better understand the boundary conditions and profiles of thinking that give rise to adaptive thinking.

Evidence was found for several profiles of both adaptive and less adaptive thinking. That is, there appears to be no single pattern of tactical adjustment that results in adaptive performance on the number-line task. Instead, several different adaptive strategies emerged that may be related to a students' initial tactic on the 0-100 line and their level of explicit conceptual number-line knowledge. As a result, it may be useful to conceptualize adaptivity in terms of a multi-dimensional space. Conceptualizing adaptivity as multi-dimensional may have implications for understanding how children with atypical development might adapt and adjust within a task space.

For example, in pilot work of the number-line estimation task with 10 elementary-aged children with ASD, two children had number-line estimation patterns that diverged substantially from the typical students. In two exploratory sessions including a number of mathematics tasks such as estimation, inversion, addition, counting, and number-line estimation, children used pencil and paper to estimate 26 targets on a 0-100 line in one session and 22 targets on a 0-1000 line in another session.

One girl with ASD, age eight years, 11 months, almost always used the same procedure—counting by ones from the 0-end on all trials on 0-100. Her pattern of estimates on the 0-100 number line was highly linear. On the 0-1000 line she initially counted from 0-end until one trial required counting to nearly 100—at which point she appeared to guess for that trial and then for subsequent trials thereafter. Her pattern of estimates on the 0-1000 line was best fit by a logarithmic equation. In contrast, a boy with ASD, age six years two months, was observed using a proportional landmark procedure on every trial. On the 0-1000 range he transformed the line to 0-100 and transformed all targets into percentages. He had patterns of estimates fit by highly linear equations on both ranges. He immediately anchored to the exact proportion on the task (apparently not using any additional anchors). His pattern of procedures across targets is best classified as the *Pure Proportions* tactic on both ranges.

Of note is that both students with ASD could be classified using the tactical ladder though their approaches to the number-task represented two extremes. The girl was using a less advanced Ends Only tactic across both ranges; the boy used the unusually advanced Pure Proportions tactic on both ranges. No typically developing student was ever observed using a Pure Proportions tactic. Like the girl with ASD, two Grade 2 students with typical development (ages 8 years and 7 years, 8 months) used an Ends Only tactic on both 0-100 and 0-1000 lines but in both cases, those students always used a qualitative adjustment from either end. In contrast, the girl with ASD often adjusted by counting. These two brief cases of students with ASD compared to the range of responses used by typically developing children, illustrate how mapping a task space and typical procedures and tactics within that task space might be used to better characterize both typical and atypical cognitive development.

Limitations

One limitation to this study is that the ranges were always presented in the same order from easy to hard, the 0-100 line followed by the 0-1000 line. This may be a problem because estimating on one range before another could induce certain kind of patterns or knowledge from the outset. A goal of this study, however, was to look at systematic changes in children's estimations as the range changed; therefore the consistent order of the ranges was necessary to compare results across all children. Also, response latencies were not recorded while students were estimating. Response latencies may have helped to triangulate claims about the kinds of solution procedures students used. Given, however, that Ashcraft and Moore (2012) found that their latency results paralleled their accuracy data, the conclusions from the present study about students' procedures and tactics are unlikely to change substantially. Response latencies might also have served as an extra dimension to use in considering how effective students were in executing their chosen solution procedures and tactics. Latency data probably would have shed more light on the adaptivity of students tactical adjustments from one range to another and may have helped create a dimension for ordering tactical adjustment from least to most efficient on the basis of accuracy and efficiency rather than just accuracy alone. A final limitation to this work is that the ranges were fairly easy so the variability in students' estimates was limited and, as a result, the kinds of procedures and tactics identified may not fully capture the range of students' thinking. Nonetheless, the task analysis, combined with the detailed and systematic coding of solution processes into 41 specific procedures, may be used to identify additional procedures and, potentially, additional tactics or sub-layers of the tactical ladder.

Conclusions

Until now, most researchers have speculated about the development of number-line estimation based on fitting equations to patterns of children's estimates. In the present study, analysis of students' trial-by-trial self-reports revealed striking variability in the processes and procedures students used to estimate. Patterns of students' selection of procedures as a function of number-line target were analyzed and six tactics were identified and organized into a developmentally tractable sequence. Rather than the development of number-line estimation being characterized as an abrupt shift from a reliance on an immature to a mature representation, the results of this study support the idea of the development of number-line estimation as a process of differentiation and hierarchical integration.

With experience students differentiate parts of the line to anchor and then integrate across the anchors systematically, from one end of the line, to the whole line, to half the line or a quarter or a tenth of the line, to construct an estimation tactic that becomes more refined over time. Moreover, students with more explicit knowledge about for example, equal intervals, scale changes, and proportions, used more advanced tactics and scored higher on a math achievement measure.

By defining solution operations at multiple levels in the context of the number-line estimation task, the opportunity to consider how children adjusted their problem-solving solutions was possible. Results revealed that rather than a single tactical shift or a particular profile of number-line estimation being the most adaptive, there was marked diversity in the kinds of tactical adjustments that led to a student being adaptive. The methodological approach used in this study lays a foundation to further investigate how students might estimate on atypical or unfamiliar number-line ranges or how children with atypical development might adapt and adjust within a task space.

Mapping the space of number-line estimation sets the stage for future studies of how children come to construct or recruit certain anchors, and learn to adjust appropriately. The taxonomy of solution operations could be used to facilitate identifying processes, procedures, tactics, and overall strategies on a variety of tasks. For example, it might be productive to identify profiles of adaptive thinking to better understand the mechanisms of adaptive tactical adjustments and investigate whether adaptivity is task specific, domain specific, or global. Knowing under which conditions children adjust successfully may guide the creation of optimal learning environments to support both typically developing children and children who have difficulty with adjusting adaptively.

References

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25*, 468-523.

Ash, K. (2012, February 8). Rethinking testing in the age of the iPad. *Education Week Digital Directions*, Retrieved from http://www.edweek.org/dd/articles/2012/02/08/02mobile.h05.html

Ashcraft, M. H. & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology, 111*, 246-267. doi: 10.1016/j.jecp.2011.08.005

Baldwin, J. M. (1898). On selective thinking. *The Psychological Review*, *5*, 1–24.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Developmental Science*, *14*, 125–135. doi:10.1111/j.1467-7687.2010.00962.x

Barth, H., Starr, A., & Sullivan, J. (2009). Children's mappings of large number words to numerosities. *Cognitive Development, 24*, 248-264. doi: 10.1016/j.cogdev.2009.04.001

Benton, B. K. (2012). *The iPad as an instructional tool: An examination of teacher implementation experience.* (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses database. (UMI No. 3516595).

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology, 46*, 545-551. doi:10.1037/a0017887

Bisanz, J., & LeFevre, J.-A. (1992). Understanding elementary mathematics. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 113-136). Amsterdam: North Holland, Elsevier Science Publishers.

Bisanz, J., Watchorn, R., Piatt, C. & Sherman, J. (2009). On understanding the principle of inversion. *Mathematical Thinking & Learning, 11,*10-24.

Bonnington, C. (2012, January 20). iPad a solid education tool, study reports. *Wired*. Retrieved from http://www.wired.com

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology, 37*, 247-253. doi: 10.1016/j.cedpsych.2012.07.001

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation *Developmental Psychology*, *42*, 189–201. doi:10.1037/0012-1649.41.6.189

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arthimetic learning. *Child Development*, *79*, 1016-1031. doi: 10.1111/j.1467-8624.2008.01173.x

Bouwmeester, S. & Verkoeijen, P. J. L. (2012). Multiple representations in number line estimation: Developmental shift or classes of representations? Cognition and Instruction, 30, 246-260. doi:10.1080/07370008.2012.689384

Brouillette, R. M., Foil, H., Fontenot, S., Correro, A., Allen, R., Martin, C. K., …, Keller, J. N. (2013). Feasibility, reliability, and validity of a smartphone based application for the assessment of cognitive function in the elderly. *PLOSone, 8*, doi: 10.1371/journal.pone.0065925

Cameron, A., & Bush, M. H. (2011). *Digital course materials: A case study of the Apple iPad in the academic environment.* (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses database. (UMI No. 3457458).

Castellan, Jr., N. J. (1991). Computers and computing in psychology: Twenty years of progress and still a bright future. *Behavior Research Methods, Instruments, & Computers, 23*, 106-108.

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences Early Edition.* doi:10.1073/pnas.1323628111

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in western and Amazonian indigene cultures. *Science, 320*, 1217-1220. doi: 10.1126/science.1156540

De Smedt, B., Holloway, I. D., & Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. *NeuroImage*, *57*, 771–781. doi:10.1016/j.neuroimage.2010.12.037

Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario F.X., …Grainger, J. (2011). Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. *PLOSone, 6*, 1-3.

Duffy, V. A. (2012). *Musical social stories and the preschool child with autism spectrum disorder.* (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses database. (UMI No. 3504191).

Ebersbach M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarty with numbers in 5- to 9-year old children: Evidence for segmented linear model, *Journal of Experimental Child Psychology, 99*, 1-17. doi:10.1016/j.jecp.2007.08.006

Ebersbach, M., Luwel, K. & Verschaffel, L. (2013). Comparing apples and pears in studies on magnitude estimations. *Frontiers in Psychology, 4*, 332. doi:10.3389/fpsyg.2013.00332

Farrington-Flint, L., Vanuxem-Cotterill, S., & Stiller, J. (2010). Patterns of problem-solving in children's literacy and arithmetic. *British Journal of Developmental Psychology*, *27*, 815–834. doi:10.1348/026151008X383148

Ford, L., Swart, S., Negreiros, J., Lacroix, S., & McGrew, K. (2010). *Use of the Woodcock-Johnson III NU Tests of Cognitive Abilities and Tests of Achievement with Canadian Populations*. Woodcock-Johnson III Assessment Service Bulletin No.12 (pp. 1–24). Rolling Meadows, IL: Riverside Publishing.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods *Psychological Bulletin*, *137*(2), 316–344. doi: 10.1037/a0021663

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

Gibson, J. J. (1986). *The ecological approach to visual perception.* Hillsdale, NJ: Lawrence Earlbaum Associates. (Original work published 1979)

Goldin-Meadow, S. (2006). Nonverbal communication: The hand’s role in talking and thinking. In W. Damon & R.M. Lerner (Series Eds.), D. Kuhn & R.Siegler (Vol.. Eds.), *Handbook of child psychology: Vol 2. Cognition, perception and language (*6th ed., pp. 336-369). New York: John Wiley & Sons.

Harrell, A. (2010, August 11). iHelp for autism. *SFWeekly News*. Retrieved from http://www.sfweekly.com

Hecht, S. A. & Vagi, K. J. (2010). Sources of group and individual differences in emerging fraction skills. *Journal of Educational Psychology, 102,* 843-859. doi: 10.1037/a0019824

Heeffer, A. (2011). Historical objections against the number line. *Science & Education, 20*, 863-880. doi: 10.1007/s11191-011-9349-0

Hiebert, J. & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: an introductory analysis. In J. Hiebert (Ed.) *Conceptual and Procedural Knowledge: The Case of Mathematics* (pp.1-27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hollands, J. G., & Dyre, B. P. (2000). Proportion judgements: The cyclical power model. *Psychological Review, 107*, 500-524. doi: 10.1037/0033-295X.107.3.500

Holt, R. (2013, March 28). Tablets to outsell PCs for the first time this year. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/technology/news/9958851/Tablets-to-outsell-PCs-for-the-first-time-this-year.html

Izard, V., Pica, P., Spelke, E. S., & Dehaene, S. (2011). Flexible intuitions of Euclidean geometry in an Amazonian indigene group. *Proceedings of the National Academy of Sciences Early Edition*. doi: 10.1073/pnas.1016686108

Jowett, E. L., Morre, D. W., & Anderson, A. (2012). Using an iPad-based video modelling package to teach numeracy skills to a child with autism spectrum disorder. *Developmental Neurorehabiliation, 15*, 304-312.

Kagohara, D. M., Sigafoos, J., Achmadi, D., O’Reilly, M., & Lancioni, G. (2012). Teaching children with autism spectrum disorders to check the spelling of words. *Research in Autism Spectrum Disorders, 6,* 304-310.

Klahr, D. & Chen, Z. (2011). Finding one's place in transfer space. *Child Development Perspectives, 5*, 196-204. doi: 10.1111/j.1750-8606.2011.0017.x

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*, 1723-1743.

Lerner, R. M. (2006). Developmental science, developmental systems, and contemporary theories of human development. In W. Damon & R.M. Lerner (Series Eds.), R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp.1-17). New York: John Wiley & Sons.

LeFevre, J.-A., Fast, L., Skwarchuk, S-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D. & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development, 81*, 1753-1767.

LeFevre, J.-A., Greenham, S. L., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction, 11*, 95-132. doi: 10.1207/s1532690xci1102_1

LeFevre, J.-A., Lira, C. J., Cankaya, O., Kamawar, D., & Skwarchuk, S-L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00641

Lemaire, P., Lecacheur, M., & Farioli, F. (2000). Children's strategy use in computational estimation. *Canadian Journal of Experimental Psychology, 54*, 141-148.

Lemaire, P., & Lecacheur, M. (2002). Children's strategies in computational estimation. *Journal of Experimental Child Psychology*, *82*, 281–304. doi: 0.1016/S0022-0965(02)00107-8

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, 124, 83-97.

Lipton, J. & Spelke, E. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76,* 978-988.

Louis Lee, N. Y. & Johnson-Laird, P. N. (2012). Strategic change in problem solving. *Journal of Cognitive Psychology.* doi: 10.1080/20445911.2012.719021

Luwel, K., Siegler, R. S., & Verschaffel, L. (2008). A microgenetic study of insightful problem solving. *Journal of Experimental Child Psychology*, *99*, 210–232. doi: 10.1016/j.jecp.2007.08.002

Lynch, J. & Redpath, T. (2012). 'Smart' technologies in early years literacy education: A meta-narrative of paradigmatic tensions in iPad use in an Australian preparatory classroom. *Journal of Early Childhood Literacy*. doi: 10.1177/1468798412453150

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Woodcock-Johnson III Normative Update: Technical Manual*. Rolling Meadows, IL: Riverside Publishing.

Moeller, K., Pixner, S., Kaufmann, Li., & Nuerk, C. (2009). *Journal of Experimental Child Psychology,103*, 503-515. doi: 10.1016/j.jecp.2009.02.006

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W.G. Chase (Ed.), *Visual information processing* (pp.283-308). New York: Academic Press.

Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Newman, R. S., & Berger, C. F. (1984). Children's numerical estimation: Flexibility in the use of counting *Journal of Educational Psychology*, *76*, 55-64.

Noël, M-P., Rousselle, L., & Mussolin, C. (2005). Magnitude representation in children. In J. Campbell (Ed.), Handbook of mathematical cognition (pp.179-195). New York: Psychology Press.

J O'Hagan. (2011, August 14). Are we screwing up the iPad? [Web log post]. Retrieved from http://edreach.us/2011/08/14/are-we-screwing-up-the-ipad/

Opfer, J. E. (n.d.). Analyzing the number-line task: A tutorial. Retrieved from: http://www.psy.cmu.edu:16080/~siegler/SiegOpfer03Tut.pdf

Opfer, J. E., & Siegler, R.S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*, 169-195. doi: 10.1016/j.cogpsych.2006.09.002

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

Pellicano, E., Aagten-Murphy, D., Attucci, C., Klaric, E., & Burr, D. (2011, May). *Number sense in autism*. Poster presented at the annual International Meeting for Autism Research, San Diego, CA.

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction*, *7*, 55–78.

Piatt, C., Volden, J., & Bisanz, J. (2011, May). An exploration using children's reasoning about math to identify cognitive profiles in autism spectrum disorder (ASD). Poster presentation at the International Meeting for Autism Research, San Diego, CA.

Piatt, C., Coret, M., Choi, M., Volden, J., & Bisanz, J. (2014). Comparing children's performance and preference for a number-line estimation task: Tablet versus paper and pencil. Manuscript in preparation.

Prather, R. W. & Alibali, M. W. (2009). Development of arithmetic principle knowledge: How do we know what learners know? *Developmental Review, 29*, 221-248.

Pressley, M., & Hilden, K. (2006). Cognitive strategies. In W. Damon & R.M. Lerner (Series Eds.), D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 511-556). New York: John Wiley & Sons.

ME Rich. (2010, April 13). Why the iPad absolutely matters [Web log post]. Retrieved from http://www.zdnet.com/blog/education/why-the-ipad-absolutely-matters/3800

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, *7*, 221-237.

Rittle-Johnson, B., & Siegler, R. S. (1999). Learning to spell: Variability, choice, and change in children's strategy use. *Child Development*, *70*, 332–348.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93,* 346-362. doi: 10.1037/0022-0663.93.2.346

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representations of the mathematical number line. *Developmental Science.* doi: 10.1111/desc.12166

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition, 17*, 759-769.

Rutherford, K. S. (1993). *Guide to analysis of language transcripts* (2nd ed.). Thinking Publication: Eau Claire, WI.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*, 418-431. doi: 10.1016/j.jecp.2012.20.012

Schaie, K. W., & Hofer S. M. (2001). Longitudinal studies in research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 53-77). San Diego, CA: Academic Press.

Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in grades 5 and 6. *Journal of Educational Psychology, 101*, 359-372. doi: 10.1037/a0013840

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., Desmedt, B., Verschaffel, L., Jacobs, A., et al. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development*, *23*, 409–422. doi: 10.1016/j.cogdev.2008.07.002

Schneider, M. & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1227-1238.

Segal, A. (2011). *Do gestural interfaces promote thinking?* (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses database. (UMI No. 3453956).

Shepard, I. J., & Reeves, B. (2011, March). iPad or iFad - The reality of a paperless classroom. Paper session presented at Mobility Conference, Abilene, TX.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250–264.

Siegler, R. S. (2005). Children’s learning. *American Psychologist*, *60*, 769–778.

Siegler, R. S. (2006). Microgenetic analysis of learning. In W. Damon & R.M. Lerner (Series Eds.), D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 464-510). New York: John Wiley & Sons.

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

Siegler, R. S., & Booth, J. L. (2005). Development of numerical estimation: A review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 197–212). Psychology Press.

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

Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives, 8,* 144-150.

Siegler, R. S., & Jenkins, E. (1989). *How Children Discover New Strategies*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science*, *17*, 237–243.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games--but not circular ones--improves low-income preschooler's numerical understanding. *Journal of Educational Psychology, 101*, 545-560.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*, 377–397.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3,* 143-150.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*, 268–288.

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General, 142*, 193-208. doi:10.1037/a0028560

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human perception and Performance, 16*, 683-692.

Spinillo, A. G. & Bryant, P. (1991). Children's proportional judgements: the importance of "half." *Child Development, 62*, 427-440.

Statistics Canada. (2006). *2006 Census of Canada topic-based tabulations, highest certificate, diploma or degree, age groups, sex for the population 15 years and over of Canada, provinces, territories, census divisions and census subdivisions, 2006-20% sample data, 2006 census of population.* (Catalogue number 97-560-XCB2006008). Retrieved June 16, 2012 from Statistics Canada.

Statistics Canada. (2006). *2006 Census of Canada topic-based tabulations, household income and household size for the private households of Canada, provinces, territories, census divisions and census subdivisions, 2005-20% sample data, 2006 census of population.* (Catalogue number 97-563-XCB2006052). Retrieved June 16, 2012 from Statistics Canada.

Stokes, T. F. & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349-367.

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*, 557–563. doi: 10.3758/s13423-011-0081-1

Sung, E., & Mayer, R. E. (2013). Online multimedia learning with mobile devices and desktop computers: An experimental test of Clark's methods-not-media hypothesis. *Computers in Human Behavior, 29*, 639-647.

Thelen, E., & Smith, L. B. (2006). Dynamic systems theories. In W. Damon & R.M. Lerner (Series Eds.), R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 258-312). New York: John Wiley & Sons.

Thompson, C. A., & Opfer, J. (2008). Costs and benefits of representational change: Effects of context on age and sex differences in symbolic magnitude estimation. *Journal of Experimental Child Psychology, 101*, 20-51.

Thompson, C. A. & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representation change in numerical cognition. *Child Development, 8*1, 1768-1786. doi: 10.1111/j.1467-8624.2010.01509.x

Thompson, C. A., & Siegler, R. S. (2010). Linear numerical magnitude representations aid children's memory for numbers. *Psychological Science, 21*, 1274-1281.

Träff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children. *Journal of Experimental Child Psychology, 116*, 139-156. doi: 10.1016/j.jecp.2013.04.007

Voutsina, C. (2012). Procedural and conceptual changes in young children's problem solving. *Educational Studies in Mathematics, 79,* 193-214. doi: 10.1007/s10649-011-9334-1

Werner, H. (1948). *Comparative psychology of mental development*. New York: International Universities.

White, S. L. J. & Szűcs, D. (2012). Representational change and strategy use in children's number line estimation during the first years of primary school. *Behavioral and Brain Functions, 8*:1. doi: 10.1186/1744-9081-8-1

Wolfe, C. R. (2006). SCiP at 35: An idiosyncratic history of the Society for Computers in Psychology, *Behavior Research Methods, 38*, 245-2

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K., Schrank, F. A., & Mather, N. (2007). *Woodcock-Johnson III Normative Update*. Rolling Meadows, IL: Riverside Publishing.

Young, C. J., & Opfer, J. E. (2011). Psychophysics of numerical representation: A unified approach to single- and multi-digit magnitude estimation. *Journal of Psychology, 2011*, 219, 58-63. doi: 10.1027/2151-2604/a000047

Appendix A

Number-line Processes

Table A1

Specific Processes: Reformulations

Specific Processes	Description
Target	Reformulate target by truncation, rounding or changing form, e.g., target is 52 but refer to as 50 or half, or target is 245 but refer to as 200
Scale [a]	By truncation, rounding or changing form, e.g., scale is 0 to 100 but refer to as 0 to 10, or scale is 0 to 531 but refer to as 0 to 500

[a] On 0-1000, counting by 100s but saying “10, 20, 30, 40” is an implicit scale change.

Table A2

Specific Processes: Anchors

Specific Processes	Description
0	Reference 0 endpoint
End	Reference endpoint (e.g., 100 or 531)
Middle	Reference the middle, endpoint/2, ½, halfway, including the physical middle even if say wrong number (note in “wrong mid” notes)
Quarter	Reference ¼ or ¾ or split in fourths (divide into 4) “25, 50,75”; counted as anchor after *Constructed Proportion* only if refer to proportion (e.g. “3/4 here”)
Other proportion	Reference other proportion, e.g., third, fifth, or tenth, or splits in thirds (“33,66”); can anchor to after *Construct Proportion* if say proportion (e.g., “1/10th here”)
Previous Trial	Reference a previous trial: “remember, where 45 was last time”
Constructed Proportion	Construct anchor by counting at least 2 proportions between 2 anchors including “mid of mid” or “half of half”
Counted	Construct anchor by counting at least two sequential numbers, “10, 20, 30”
Non-proportional Construct	Anchor constructed from increment or decrement move: computation, adjustment, measurement, or between X and Y, basically - you can tell where the anchor comes from in relation to other anchors
Other	Other anchors, including vague references to an area (“in the 30s”)

Table A3

Specific Processes: Adjustments

Specific Processes	Description
Qualitative	Adjust qualitatively ("up/down"; "a bit"; "some"; "near"; "far from"; "close to", "left", "right", "not", "almost", "about")
Quantitative	Adjust quantitatively ("add/subtract"; "about x amount"; "x mm"; "greater or less than", "a few ticks over")
Count by Ones	Counting ("up 1, 2, 3, 4") two or more sequential numbers
Count by Tens	Counting ("60, 70") two or more sequential numbers
Count Other	Counting ("3, 6, 9") two or more sequential numbers
Middle	Finding "middle", "halfway between X and Y" or "between X and Y"
Proportion	Adjust proportionally ("quarter of the way from X", "third of way from X")
Previous Trial	Adjust using amount from a previous trial
Guess	Adjust by guessing ("then I guessed", "somewhere around here", "maybe here", "around") or a back-and-forth gesture

Appendix B

Comparability of Silent vs. Self-Report Estimates

In the last decade, research on children's number-line estimation has been plentiful. Number-line tasks have been used to measure how children represent numbers and how those representations change over time (Barth & Paladino, 2011; Booth & Siegler, 2006; Siegler & Opfer, 2003; Siegler, Thompson, & Opfer, 2009; Slusser et al., 2013). Studies of performance on number-line estimation are important not only for understanding the development and organization of number-line representation, but also because estimation is an important everyday mathematical skill. Performance on number lines is highly correlated with later basic math skills such as arithmetic (Booth & Siegler, 2008). One way to gain insight into the development of children's thinking about number-line estimation is to ask them about their thinking while estimating (Schneider et al., 2008).

Data from think-aloud protocols have revealed new insights into children's thinking about other kinds of mathematical problems such as addition and inversion. Asking students to think-aloud on every problem however, may lead to increased awareness of the task and cause children to react or behave differently—a phenomenon called reactivity. Reactivity effects have been observed specifically when participants are asked to explain or reflect on their thinking (Crowley & Siegler, 1999; Fox, Ericsson, & Best, 2011; Siegler, 2002). In contrast, think-aloud procedures, defined as concurrent verbalizations or "task-relevant thoughts generated between the start of a primary task and completion of the associated task," show minimal reactivity effects (Fox, et al., 2011; Johnson, 1993). Most studies of number-line estimation focus only on children's placement of the estimate on the number line. A few early studies included some think-aloud protocols on a few trials (e.g., Newman & Berger, 1984; Pettito, 1990). To advance our understanding of how number-line estimation develops, we explored whether asking children to describe their estimations would change their performance on the task. We compared the accuracy of Grade 6 students in one of two reporting conditions, silent or think-aloud, to determine whether thinking aloud influences performance on a number-line task.

All students were silent while making their estimates on the first set of 10 targets on a 0-100 number line. Then, while making estimates for the second set of 10 targets on the same line, half the students were instructed to remain silent, and the other half were instructed to describe their estimation procedures aloud. As expected, no difference in performance on the task was found between students who estimated silently and those who described their thinking. A practice effect was found for all students, regardless of reporting condition, whereby performance for the last 10 trials was significantly better than on the first 10 trials. These results are consistent with the previous literature on think-aloud protocols and confirm that asking students to think-aloud while estimating on a number-line task results in performance that is comparable to when students estimate silently.

Method

Participants. Thirty-two Grade 6 students (16 girls) from a small suburb of a Canadian city participated. Their ages (years; months) ranged from 9;11 to 12;6 with a mean of 11;10. Students in the think-aloud condition were drawn from the dissertation research that included several numerical ranges; students in the silent condition were drawn from another methodological investigation of number-line estimation (Piatt et al., 2014). With permission from the school district and school principals, children were recruited through classroom teachers. Testing took place in a quiet room at their school. The Human Research Ethics Board at the University of Alberta approved the research. Two graduate and one undergraduate psychology student(s) served as experimenters.

Number-line estimation task. The task was adapted from the number-to-position estimation procedures of Booth and Siegler (2006) and Barth and Paladino (2011). For the 0-100 range, two targets were selected from every decade. Stimuli were presented as two sets of 10 targets with a short break in-between (see Table B1). All students were silent for the first 10 trials and saw stimuli from Set 1. Next all participants saw stimuli from Set 2; half of the participants continued to estimate silently for the last 10 trials (silent condition) whereas the other half reported their thinking aloud on the last 10 trials (think-aloud condition). Within each condition, participants were assigned to one of two semi-random presentation orders of targets, Orders A and B (Table B1). Presentation order was also counterbalanced across gender.

Table B1

Stimuli for Number-line Estimation Task

Condition Order	Set Order	Target Values: Order A[a]	Target Values: Order B [a]
Silent (All students)	Set 1	21, 48, 61, 94, 57, 36, 88, 17, 8, 76	76, 8, 17, 88, 36, 57, 94, 61, 48, 21
Silent or Think-aloud	Set 2	96, 33, 4, 52, 81, 42, 14, 67, 25, 72	72, 25, 67, 14, 42, 81, 52, 4, 33, 96

[a] Targets were ordered semi-randomly with the constraint that no targets from the same decade or century appeared in adjacent positions, and that trials did not appear in ascending or descending order. The second order is the reverse of the first.

Task presentation. The number-line estimation task was presented as an application[1] (app) on a tablet (iPad). On the iPad, number-line stimuli were presented as a straight line on the touch screen, 18 cm in length, with 0 marked at the left end and the endpoint of the number-line range (e.g., 100) marked at the right end. The target number was centered 2.5 cm above the line. To make their estimates on the iPad, participants used a Kuel H10 high sensitive stylus.

Instructions for administration. Task procedures were adapted from Booth and Siegler (2006) and Barth and Paladino (2011). The task began by obtaining students' assent and orientation to the iPad and stylus. Next, students were introduced to the number-line task. The experimenter said, while showing the child a ruler and gesturing to the ends of the ruler and numbers in between, "A number line is like a ruler. It has numbers and marks at both ends and all along it." The experimenter then started the iPad,

[1] The application was developed in collaboration with a graduate student in computing science at the University of Alberta, Michael Choi.

showed the child a screen with a 0-100 line, and said, “This is a number line too, but on this number line only the numbers at the ends are written in.” A new screen appeared with the numbers on each end enlarged 200% and then scaled down to the normal size to draw attention to the number-line range. During this screen the experimenter said:

> The flashing red numbers remind us what number line we are working on. We are working on a number line from 0 to 100. Remember, tap quickly, like this *(experimenter demonstrates)*, and make sure your hand is not touching the screen. Let’s say I put the answer for the number at the top of the screen (100 shown on screen) here *(experimenter marks near but not on 100)*, but that is not where I think the answer is. It’s okay to make another mark like this *(experimenter marks on 100)*. When you tap your best answer, tell me “done”, and then I will tell you it’s okay to press the OK button to lock in your answer *(experimenter demonstrates pressing OK button above the mark)*.

Finally, students practiced using the stylus, making marks and pressing the OK button by showing the experimenter where the practice targets, 0 and 100, were on the line. “I know you can already see where the number 0 is, but to practice, tap where 0 should go on the number line.” The experimenter than said “Good! Now tap where 100 goes on the number line.”

Just prior to starting the first half of trials, the experimenter reminded the student to make his or her best answer as quickly as possible and then to tell the experimenter “done” before locking in their answer with the OK button. Students then estimated 10 targets on a 0-100 number line (Table B1). After completing the first set, a screen appeared with a video of some cheerful music and animated dancing bunnies holding a sign that said “Nice Job!”

After the first set, students were told “Good job!” and given a short break. Students in the Silent condition went on to estimate the positions of 10 more targets on a 0-100 number line (Table B1). Students in the think-aloud condition were told:

> Now we are going to do something a little different. This time I am going to ask you to think out loud and tell me in a loud and clear voice what you are doing in your head as you decide where to put numbers on the number line.

To practice thinking out loud, the student was asked to draw a house on a commercially available app (Drawing Pad) on the iPad, and told “I want you to draw a house and at the same time as you are drawing, tell me in a loud and clear voice, what you are doing as you draw.” If the student seemed confused or said that he or she did not understand, the experimenter demonstrated by drawing the base of a house and two walls and saying, “I am drawing the walls of the house,” and then she drew the roof saying, “and now I am drawing the roof. ” The experimenter erased her work and handed the stylus over to the student saying, “Now you try.” If the student started drawing but did not report on their thinking, the experimenter prompted, “Tell me what you are thinking as you draw.” All students were able to clearly report their thinking as they drew a house.

Just prior to beginning the second set of 10 targets on a 0-100 number line, all students were reminded to make their best answer as quickly as possible. Students in the think-aloud condition were also told that the experimenter, after each target, would always ask the student more about her thinking by asking “How did you figure out where to put the number?” If the student’s answer was unclear or vague, the experiment prompted, "How did you figure it out?" or "How did you put it there?" for the purpose of obtaining a self-report of solution processes (e.g., Siegler & Stern, 1998).

Data collection. On all trials accuracy was recorded using the iPad app. Additional marks, such as drawing a succession of lines to count off units on the screen, were also recorded by the app.

Results and Discussion

Students in both reporting conditions across both halves of the task had nearly perfect accuracy as illustrated by the data in Figure B1.

To address whether general accuracy on the number-line estimation task varies as a function of thinking aloud, each students' mean percent absolute error (PAE) was calculated (in Booth & Siegler, 2006, 2008):

$$PAE = \frac{|\text{Estimated Position} - \text{Target Presented}|}{\text{Numerical Range}} \times 100$$

For example, if a child was asked to estimate the location of 43 on a 0 to 100 number-line and placed the mark at the location that corresponded to 56, the percent absolute error on that trial would be 13%: (|43-56|/100) x 100.

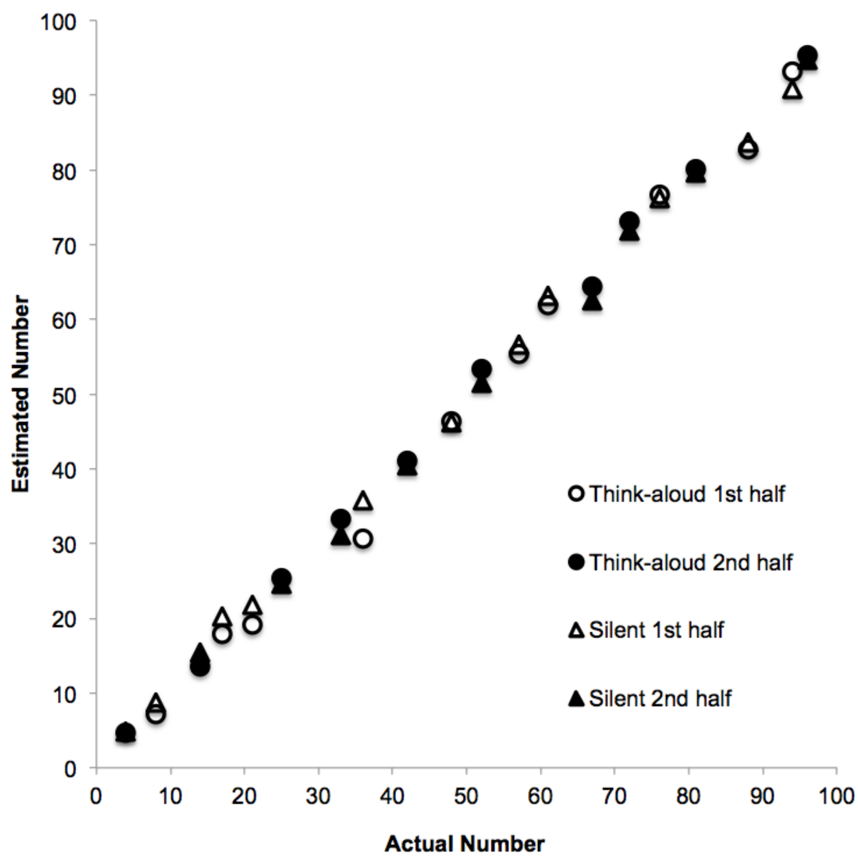


Figure B1. Grade 6 students' median number-line estimates for each half (Set 1 and 2) in one of two reporting conditions, silent and think-aloud.

A Half (First, Second) x Condition (Silent, Think-aloud) repeated measures ANOVA on students' mean PAEs revealed that performance on the second set of estimates (M = .04, [.03, .05]) was significantly better than on the first set (M = .05, [.04, .05]), $F(1, 30) = 8.38$, $MSE = .01$, $p = .01$, $\eta_p^2 = .22$; means shown with 95% confidence intervals. As expected, there was no effect of condition and no interaction, confirming that thinking aloud did not affect students' performance on the task. To determine whether thinking aloud may have influenced performance somewhat differently for a sub-set of students who were either wildly accurate or inaccurate we examined the distribution of students' PAE scores. Only one student's PAE was less than 2%, and all other students' PAEs were evenly distributed between 2% and 7%. Thus, there was no subset of students whose performance was different enough to analyze separately.

Finally, to see whether thinking aloud might influence performance on individual trials across trials, the distribution of mean percent absolute errors was graphed in Figure B2 as a function of trial. Figure B2 further demonstrates that error rates were low. All estimates, regardless of being in the first or second half, fell within a few percentage points of one another. No discernable pattern of change in performance across trials was found, nor was there found a subset of trials that might be confidently analyzed separately for reactivity effects.

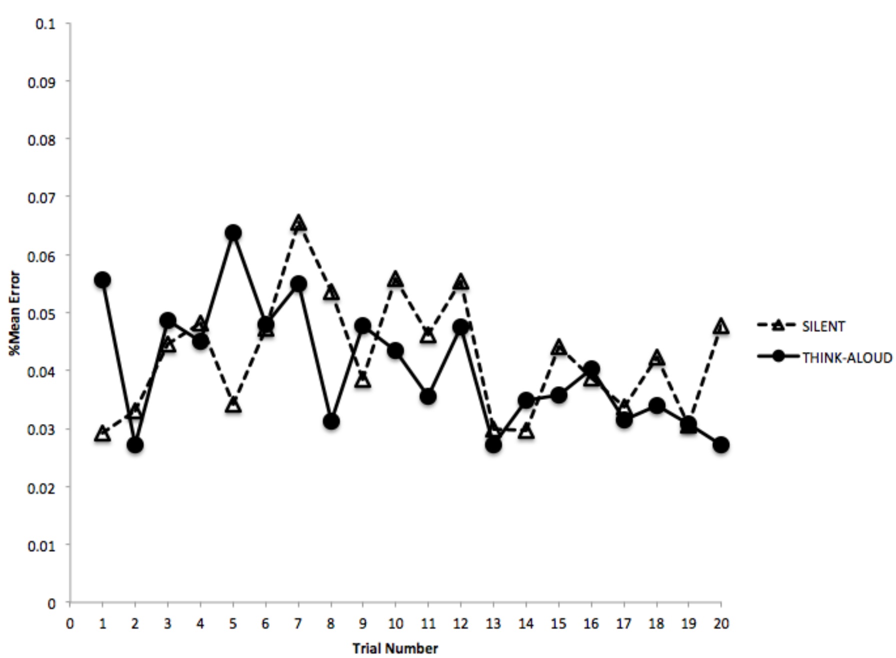


Figure B2. Mean percent absolute errors, for each reporting condition, silent and think-aloud, as a function of trial number on the 0-100.

Apart from a general practice effect where all children performed better on the second set of 10 trials than the first set, no differences were found as a function of reporting condition. In fact, in the last 10 trials, the two functions are almost identical, which again demonstrates that thinking aloud did not influence students' performance on the task. Grade 6 students performed well on the 0-100 number line and, as a result, changes in performance due to thinking aloud are difficult to detect. To address this limitation, it would be useful to investigate the effects of thinking aloud on number-line ranges with more variability in performance such as larger or more unfamiliar number-line ranges.

Appendix C

Explicit Number-Line Knowledge Task (K_{num})

Students first explained how the number line worked to an imaginary student, Casey (Table C1). These data are not coded at this time for two reasons. First, there was considerable variability in how much students said in response to the question and in the quality of their responses. Second, the targeted questions about specific properties of number lines more systematically captured the kinds of things students said in response to the first open-ended question (if they said much at all). After the first open-ended question, the experimenter asked, always in the same order, a series of questions about the number line. Questions measured students' understanding of several components of number-line knowledge: basic knowledge, equal interval, scale, and proportions (Table C1). These components helped to organize the task as well as the data file and coding scheme. A summary of how the task was coded follows Table C1. Following the summary is the detailed coding protocol. On a sample of 18% of the data, inter-rater reliability was 90%.

Table C1

Questions on Explicit Number-line Knowledge Task

Section	Questions
Open-ended	Casey is a student at another school learning about number lines. Using this number line *(give 0 to 100 number line template and pencil to student)*, how would you explain to Casey what the number line is and how it works?
Basic	1. Where are the small, medium, and large numbers on the line? 2. If we put all of the numbers on this line, how many numbers would there be? Why? 3. Where would 103 be? Why would you put it there?
Anchors and adjustments	4. Can you show me where 50 would be? Why did you decide to put it there? 5. Show me where 30 goes on the line. Show me where 60 goes on the line. How did you decide to put those numbers right there and not somewhere else on the line?
Equal interval and scale	Here is the space (*point to numbers they made*) between 30 and 50, and here is the space between 60 and 50. I am going to ask you questions about the space between numbers. Okay? 6. Does the space between 0 and 10 have to be the same as between 90 and 100, or can it be longer or shorter? Why or why not? 7. Does the space between 10 and 11 have to be the same as the space between 90 and 91, or can it be longer or shorter? Why or why not? 8. If this end point changed to be a bigger number, such as 1000, what would happen to the space between 50 and 60 *(point to space)?* Why? 9. If this end point changed to be a bigger number, such as 1000, what number would be exactly in the middle *(point to middle)*? Why?
Proportions	10. Can you show me halfway on the number line? Why did you decide to put ½ there? 11. Can you show me ¼ of the way on the number line. Why did you decide to put ¼ there? 12. Last question: can you show me where ¾ of the way is on the number line. Why did you decide to put ¾ there?

Last Questions (Version 2 only)	13. If this end point changed to be a bigger number, such as 1000, where will 50 be *(point to their 50 mark)?* Will it still be in the middle or will it be someplace else? If someplace else, where? Show me. How about 60? Where will it be if the end point changed to a bigger number such as 1000? Show me. 14. Now, this is where you put 50 and 60 *(point)* if we imagine the line is from 0 to 1000, is the space between 50 and 60 the same as it was on a 0 to 100 number line? Or is it larger or smaller? Why?

Summary of Number-line Knowledge Task Coding

Codes were developed iteratively as we anticipated some types of answers and codes, as well as discovering consistencies in other kinds of responses students gave. Coders sought to capture the information as stated by the participant. When in doubt, coders referred to the video recording of the student completing the task. Coders used photocopied versions of the student's original explanation data sheet so that coding notes and marks could be made on next to the data, thereby keeping a coding trail while still preserving the original data. For example, when it was unclear which mark to use for an estimate, coders watched the video and indicated on the photocopied sheet which of the student's marks was used. Finally, when in doubt about a code the issue was decided by group consensus. Data were coded by three undergraduate research assistants and the first author. On a sample of 18% of the data, inter-rater reliability was 90%.

Basic Knowledge. Three questions measured basic number-line knowledge:

1. Where are the small, medium, and large numbers on the line?
2. If we put all of the numbers on this line, how many numbers would there be? Why?
3. Where would 103 be? Why would you put it there?

All students answered the third question correctly. Because the third question had no discriminatory value it was excluded from further analyses. Answers to both questions were scored as correct (1) or incorrect (0). Answers of "don't know" were scored as incorrect. One student missed both basic questions. Ten students answered only one basic question correctly, and the remaining 61 students answered both questions correctly.

Equal Interval. Knowledge of equal intervals was measured with two questions:

1. Does the space between 0 and 10 have to be the same as between 90 and 100, or can it be longer or shorter? Why or why not?
2. Does the space between 10 and 11 have to be the same as the space between 90 and 91, or can it be longer or shorter? Why or why not?

For both questions, answers were correct (1) only when students reported that the answer was "the same" along with a correct explanation. All other answers were incorrect (0). Twelve students answered both questions incorrectly, and 11 students answered only one question correctly. The remaining 49 students answered both questions about equal intervals correctly.

Scale. To assess understanding of scale, two questions were asked:

1. If this end point changed to be a bigger number, such as 1000, what would happen to the space between 50 and 60 *(point to space)?* Why?
2. If this end point changed to be a bigger number, such as 1000, what number would be exactly in the middle *(point to middle)*? Why?

For the first question, only answers that referred to the space changing to be both smaller and moving back on the line were correct (1); all other answers were incorrect (0). For the second scale question, only students that stated the middle would be 500 with an acceptable explanation were scored as correct (1). All other answers were incorrect (0). Five students were not asked the second question because it was added after discovering that the first question alone did not sufficiently capture what happens to the middle number when the scale changes from 100 to 1000. Those five students all answered the first question incorrectly. Twenty-five students missed both questions about scale and 35 students missed one question about scale. Of the remaining students asked both questions only 7 answered both questions correctly.

Two additional questions were added about how estimates change when the scale changes to elucidate more clearly, what students understood about how the scale changes:

1. If this end point changed to be a bigger number, such as 1000, where will 50 be *(point to their 50 mark)?* Will it still be in the middle or will it be someplace else? If someplace else, where? Show me. How about 60? Where will it be if the end point changed to a bigger number such as 1000? Show me.
2. Now, this is where you put 50 and 60 *(point)* if we imagine the line is from 0 to 1000, is the space between 50 and 60 the same as it was on a 0 to 100 number line? Or is it larger or smaller? Why?

These two questions were not formally considered as part of scale knowledge because 13 students were not asked these questions. However, were these questions used to assess understanding they would have been used to increase the strictness of the coding answers about scale. Of the 42 students who scored a 1 or 2 on the scale component, 38 students were asked the additional questions about changing scale. All but 1 of these 38 students said that 50 would be someplace else on the line if the scale changed to 1000, and all but 1 made their estimates of 50 and 60 on a 1000 scale below the halfway mark. This consistency between coding of scale knowledge and the additional knowledge from having asked two more questions suggest that overall, the original questions captured what students understood fairly well. In both cases, it was the same student (ID 219) who seemed to show somewhat inconsistent knowledge. Student 219 answered the second scale question correctly despite not showing the expected pattern on the two additional questions. To maintain consistency in the coding his score on the scale questions was not changed.

Proportions. Students' understanding of proportion on 0-100 was assessed by considering their answers and estimations of each of three proportions:

> Can you show me ½ (¼, ¾) ("half", "one-quarter", and "three-quarters" way on the number line? Why did you decide to put ½ (¼, ¾) there?

Answers for half (½) were correct (1) when three criteria were met: (a) the student's estimates of 50 were between 45 and 55, (b) the student made consistent (off by less than 1 number) estimates for two estimates of half (one estimate came earlier in the task), and (c) the student made an explicit reference (verbally or with a gesture) to 50, the middle, or showed how to find half by using spaces to split the line into two halves at least once. Answers for quarter (¼) were correct (1) when two criteria were met: (a) the student made an explicit reference (verbally or with a gesture) to 25, or half of half, and (b) the student's actual estimate of 25 was between 20 and 30. Answers for three quarters (¾) were correct (1) when two criteria were met: (a) the student made an explicit reference (verbally or with a gesture) to 75, or half of half (indicating halfway between 50 and 100), or demonstrated using three ¼ spaces to derive ¾, and (b) the student's actual estimate of 75 was between 70 and 80. All other answers were incorrect (0). Sixteen students missed all proportions, 28 students showed understanding of only one proportion, 13 students had two of three proportions and 15 students understood all three proportions.

Index of number-line knowledge. Student's knowledge of number lines was indexed by summing across the 9 items described above to reveal their K_{num} score. K_{num} scores could range from 0 (zeros on all nine questions) to 9 (ones on all questions). The lowest K_{num} score was a 1 and the highest a 9. The mean value for K_{num} was 4.92 and the median was 5.33. The five students who were not asked the second scale question are included but their score, out of eight, was weighted by multiplying their score by 1.125 to get a score comparable to the nine-point scale.

Number-line Knowledge Task Coding Protocol

Participant Identification. Information for each participant is in the first four columns and includes ID number, grade, age in months, and the version of the explanation task each student was given.

Column A: Participant ID number in "ID_Num"

First digit = Grade (2,4,6)

Second two digits = identifier within grade

Column B: Grade in "Grade"

2 = Grade 2

4 = Grade 4

6 = Grade 6

Column C: Participant's age in months in "Age_Mns"

Note. 110 = 110 months. Age originally calculated in years and months by subtracting student's birthdate (supplied by parents in consent form) from date of first testing session.

Column D: Version of task in "Ver"

1 = Version 1

2 = Version 2

Note. Version 1 was given to 13 of 72 finalized students. Version 1 did not include Questions 13 and 14 and estimates of 50 and 60 on 1000 because after starting to give the task we realized we needed to be more explicit in asking about questions that clarified whether students understood that the middle changed when the scale changed. Of these 13 students, the very first students (n = 5) were also not asked Question 9. Version 2 includes all questions from Version 1 as well as Questions 13 and 14.

Basics. Students first answered several questions about the general structure of the 0-100 number line.

Question 1. *Where are the small numbers on the line? Where are the large numbers on the line? Where are the medium numbers on the line?*

Column E: Codes for Question 1 in "Q1"

0 = Incorrect

1 = Correct

2 = Don't Know

Note. Correct response must include being accurate on all three areas: "small around 0, medium around middle, large around 100".

Question 2. *If we put all of the numbers on this line, how many numbers would there be? Why?*

Column F: Codes for Question 2 in "Q2"

99/100/101 = Number student stated

2 = Don't Know

3 = Other Answer

Note. Only the first part of this question is coded here. The responses given for why were generally redundant with the number students gave. For example "100 because the line goes up to 100." Typical numbers students gave: 99, 100, 101 but also heard 145. Don't Know includes responses such as "I don't know", "not sure". Other Answer refers to answers that do not include an exact number. For example one student said "a lot because can go up to more than 1 million".

Question 3. *Where would 103 be? Why would you put it there?*

Column G: Codes for Question 3 in "Q3"

0 = Incorrect

1 = Correct

2 = Don't Know

Note. Only the first part of this question is coded here. All students answered this question correctly and their reasoning was "100, 101, 102, 103" or "103 is after 100" or "100 and three more numbers".

Anchors. Next, students answered several questions about anchors on the 0-100 number line: All estimates were calculated by measuring the student's response with a ruler and calculating student's exact estimates by measuring the student's mark from the

0 end of the line, dividing that number by the total length of the line and multiplying that value by the given endpoint of the line, in this case, 100 (based on Opfer, nd.). When unclear about a student's mark, coders watched the video. The mark used to assess the student's answer is marked in the photocopied version of the coding sheets. This method was used for all marks that students made as estimates on the explicit number-line knowledge task.

Question 4. *Can you show me where 50 would be? Why did you decide to put it there?*

Column H: Student's estimate for 50 "Est50"

Question 5. *Show me where 30 goes on the line. Show me where 60 goes on the line. How did you decide to put those numbers right there and not somewhere else on the line?*

Column I: Student's estimate for 30 "Est30"

Column J: Student's estimate for 60 "Est60"

Questions 4 and 5 about why student's made their marks where they did was not coded at this time because we relied only on their actual estimate as a way to make inferences about their ability to anchor along the line.

Equal Intervals. Next, came questions designed to elucidate whether the student understood the nature of equal intervals on the number line. The experimenter first said, "Here is the space between 30 and 50, and here is the space between 60 and 50. I am going to ask you questions about the space between numbers. Okay?" Then the experimenter asked:

Question 6. *Does the space between 0 and 10 have to be the same as between 90 and 100, or can it be longer and shorter? Why or why not?*

Column K: Code for Question 6 "Q6"

Question 7. *Does the space between 10 and 11 have to be the same as between 90 and 91, or can it be longer and shorter? Why or why not?*

Column L: Code for Question 7 "Q7"

For both Q6 and Q7:

1 = Same because:

"same amount of numbers" or "same amount", "look the same", "both 10 off" or "both 1 away", or "same distance"

2 = Longer, shorter, longer or shorter, not the same with explanation

3 = Don't Know or "no, I don't know", "not sure", "I forget"

Scale. Students' explanations about what happens to certain numbers when the scale changes from 0-100 to 0-1000.

Question 8. *If this endpoint changed to be a bigger number, such as 1000, what would happen to the space between 50 and 60 (point to space)? Why?*

Column L: Code for Question 8 “Q8”

1 = Shorter/smaller/closer and move down/back/near 0 (correct)

2 = Move down/back/near 0 only

3 = Shorter/smaller/closer only

4 = Same AND move down/back

5 = Same/no change only

6 = Would add 0s to the numbers 50 and 60

7 = Longer/wider/farther apart only OR Move up

8 = Conflicting (two of the above combined) or Don’t Know

Question 9. *If this end point changed to be a bigger number, such as 1000, what number would be exactly in the middle (point to middle)? Why? (Note: this question was added to Version 1.1 and 2 so 5 students were not given this question at all.)*

Column N: Code for Question 9 “Q9”

1 = 500 because “500 is half of 1000”, “50 with a 0”, “500 + 500 = 1000”

2 = 100 because “100 is half of 1000” or “don’t know why”

3 = 550 because “550 is half of 1000”

4 = Don’t Know “not sure” or “I don’t know”

5 = Other incorrect answer and explanation

999 = question not asked (student given Version 1)

Proportions. All students were asked about three proportions on the number line.

Question 10 *Can you show me ½ way on the number line? Why did you decide to put ½ there?*

Column O: Codes for "Half_is_fifty"

0 = reference another number (verbally, gesture or label) ≠ 50

1 = reference 50 verbally or with gesture

2 = Don't Know

3 = Unclear answer

4 = Reference Middle

5 = Derive half estimate by using spaces (but not explicitly 50)

6 = Other

Note. The 1 code is clearly correct and 0 is clearly incorrect. Codes 4 and 5 are "somewhat correct" but do not meet our strict criteria of half = 50.

Column P: Codes for "Half_equals"

50 = correct value given

= other number student gave

-222 = no value given

Note. No question was explicitly asked to get at this so some students did not answer (-222) but otherwise this captures what value a student thinks equals half on the 100 range.

Column Q: Codes for "Locate_half"

Student's exact estimate

Question 11. *Can you show me ¼ way on the number line? Why did you decide to put 1/4 there?*

Column R: Codes for "Quarter_is_twentyfive"

0 = reference another number (verbally, gesture or label) ≠ 25

1 = reference 25 verbally or with gesture

2 = Don't Know

3 = Reference Half of Half

4 = Correct side of 50 and correct order of ¼ and ¾, still incorrect such as estimating above 30 or above 25)

5 = Wrong side of 50 and correct order of ¼ and ¾

6 = wrong side of 50 and wrong order of ¼ and ¾

7 = Vague or unsure (e.g., two estimates or between two points)

Column S: Codes for "Quarter_equals"

25 = correct value given

= other number student gave

-222 = no value given

Column T: Codes for "Locate_quarter"

Student's exact estimate

Question 12. *Can you show me ¾ way on the number line? Why did you decide to put ¾ there?*

Column U: Codes for "Threequarters_is_seventyfive"

0 = reference another number (verbally, gesture or label) ≠ 75

1 = reference 75 verbally or with gesture

2 = Don't Know

3 = Reference Half of half (indicating between 50 and 100)

4 = Correct side of 50 and correct order of ¼ and ¾, still incorrect such as estimating above or below 75)

5 = Wrong side of 50 and correct order of ¼ and ¾

6 = wrong side of 50 and wrong order of ¼ and ¾

7 = Vague or unsure (e.g., two estimates or between two points)

8 = Other that is incorrect

9 = Use three ¼ spaces (with gesture or out loud) to derive ¾ estimate but does not give number

Column V: Codes for "ThreeQuarters_equals"

75 = correct value given

= other number student gave

-222 = no value given

Column W: Codes for "Locate_threequarters"

Student's exact estimate

Additional Scale Questions. These last two questions were added to Version 2 to clarify students' understanding of scale changes.

Questions: *If this end point changed to be a bigger number, such as 1000, where will 50 be (point to their self-made 50 mark)? Will it still be in the middle or will it be someplace else? If someplace else, where? Show me.*

Column X: Codes for "Q13"

0 = middle

1 = someplace else (verbal or put someplace else)

2 = Don't Know

3 = Other

4 = Same place

5 = Something else meaningful

999 = question not asked (student given Version 1)

Column Y: Codes for "Estimate50_on1000"

Student's exact estimate

Questions: *How about 60? Where will it be if the end point changed to a bigger number such as 1000? Show me.*

Column Z: Codes for "Estimate60_on1000"

Student's exact estimate

Questions: *Now, this is where you put 50 and 60 (point to their marks). If we imagine the line is from 0 to 1000, is the space between 50 and 60 the same as on the 0-to-100 number line? Or is it larger or smaller? Why?*

Column AA: Codes for “Q14”

1 = Smaller

2 = Larger

3 = Same

4 = Don’t Know

5 = Not the same and but does not specify larger or smaller

999 = question not asked (student given Version 1).

Appendix D

Comparing Children's Performance and Preference for a Number-line Estimation Task:

Tablet Versus Paper and Pencil

Carley Piatt, Marian Coret, Michael Choi, Joanne Volden, and Jeffrey Bisanz

University of Alberta

Author Note

Carley Piatt, Department of Psychology, University of Alberta; Marian Coret, Department of Psychology, University of Alberta; Michael Choi, Department of Electrical and Computer Engineering, University of Alberta; Joanne Volden, Department of Speech Pathology and Audiology, University of Alberta; Jeffrey Bisanz, Department of Psychology, University of Alberta.

Michael Choi is now at Dynastream Innovations.

Corresponding author:

Carley Piatt

Dept. of Psychology, P-217 Bio Sci Bldg, University of Alberta, Edmonton, AB, T6G 2E9.

Email: piatt@ualberta.ca

Phone number: +17804925262 (work); +17802312779 (cell)

Abstract

Tablet computers are positioned to be powerful, innovative, effective, and motivating research tools. It behooves psychological researchers, especially in domains with educational implications such as mathematical thinking and learning, to integrate research methods with technology already gaining traction in education and other domains. In the last decade, research on children's number-line estimation has been plentiful. In fact, commercially available computer applications for research and teaching with tablet computers have become available recently for tasks such as number-line estimation. In this study we addressed two questions. First, is performance with paper and pencil comparable to performance on a tablet? Second, do students prefer either method of presentation and, if so, why? Students' performance on a number-line estimation task was comparable on the tablet to data collected using paper and pencil. Moreover, students liked both presentation conditions equally but, when given a choice, most students preferred the tablet. Students' reasons for preferring one presentation condition to the other were explored, along with implications for using tablet applications in developmental science.

Keywords: tablets, cognitive psychology, mathematical thinking, elementary students

Comparing Children's Performance and Preference for a Number-line Estimation Task: Tablet Versus Paper and Pencil

It has been reported that for the first time tablets will outsell personal computers (Holt, 2013). How research methods examining complex behaviour advance to match technology, and conducting top-notch scientific research with consumer-grade electronic equipment have always been part of the business of computing in psychology (Castellan, 1991; Wolfe, 2006). Smartphone technology and tablet computers are prime examples of the kinds of equipment Wolfe suggests psychologists should productively integrate into psychological research. Despite the commercial availability of computer applications for research and teaching with tablet computers, little is known about how conducting experiments with computer applications and tablets compares to using more traditional research methods such as pencil and paper. As tablet computers become more common it is important to gain a better sense of whether psychological studies employing tablets for task presentation and data collection are comparable to more traditional formats used to present tasks. We explored the use of a tablet computer and a specific software application (app) in an area that has been studied extensively using pencil-and-paper tasks: the development of children's mathematical thinking about number-line estimation.

Smartphone technology has already been targeted as an innovative, consumer-grade, powerful, personal, and engaging way to easily and consistently collect data from users around the world (Dufau et al., 2011; Miller, 2012). As Dufau and colleagues noted, smartphones and tablet computers are ideally adapted to study cognitive functions because they are portable, easy-to-use, multi-media enabled, identical across countries, and designed to use apps available on the Internet, making it possible to easily collect data. Dufau et al. successfully used smartphone technology to collect response latencies on a word/non-word lexical decision task programmed as an app (http://www.sciencexl.org/) from over 4,000 smartphone users. Latency data from these users were highly correlated and similarly distributed to latencies collected using traditional laboratory methods. Recently, researchers demonstrated the feasibility, reliability, and validity of using a smartphone-based application to assess cognitive functioning in the elderly (Brouillette et al., 2013). These studies illustrate the comparability of the smartphone platform to more traditional research methods.

Tablet computers are also positioned to be powerful research tools, and have the potential to be innovative, effective, and motivating educational tools (Benton, 2012; Bonnington, 2012; Lynch & Redpath, 2012; Shepard & Reeves, 2011). Tablets are suited for supporting learning based on principles of universal design and inclusiveness (O'Hagan, 2011; Rich, 2010). Tablets have also been proposed to be especially engaging and motivating for students with developmental disorders such as autism spectrum disorders (Duffy, 2012; Harrell, 2010; Jowett, Morre, & Anderson, 2012; Kagohara, Sigafoos, Achmadi, O'Reilly, & Lancioni, 2013). Kagohara and colleagues (2013) reviewed 15 studies involving tablet computers used in educational programs and interventions for individuals with atypical development such as autism spectrum disorders. They found that tablets could productively help individuals with developmental differences to enhance academic, communication, leisure, and employment skills. Finally, educators have already begun piloting ways of using tablets for efficient, effective assessment in classrooms because tablets are useful for both asking questions and getting answers, as well as for recording and capturing a process (Ash, 2012).

It behooves psychological researchers, especially in domains with educational implications such as mathematical thinking and learning, to strive to productively integrate research methods with technology already gaining traction in education (Ash, 2012; Rich, 2010). Despite interest in tablet computers as a research tool, little is known about the comparability of computer tablets to traditional research methods such as pencil and paper. In this study we used a tablet computer and designed an application (app) in an area studied extensively using pencil-and-paper tasks: number-line estimation.

Number-line tasks have been used to measure how elementary-aged children represent numbers and how those representations change over time (Barth & Paladino, 2011; Booth & Siegler, 2006; Siegler & Opfer, 2003; Siegler, Thompson, & Opfer, 2009; Slusser et al., 2012). In the number-to-position number-line estimation task, participants are shown a horizontal line with, for example, 0 at the left endpoint and 100 at the right endpoint. Participants are asked to locate a target such as 57 on the line. Studies of performance on number-line estimation are important not only for understanding the development of number-line estimation, but also because estimation is an important everyday math skill. Performance on number lines is highly correlated with basic math skills such as arithmetic (Booth & Siegler, 2008) and measures of math achievement overall (Booth & Siegler, 2006).

Investigating whether students perform similarly on number-line estimation on a tablet compared with pencil-and-paper is important because most existing research has been conducted with paper and pencil (Barth & Paladino, 2011; Booth & Siegler, 2006;

Laski & Siegler, 2007; Opfer & Thompson, 2008; Petitto, 1990; Piatt, Volden, & Bisanz, 2011; Siegler & Booth, 2005; Siegler & Opfer, 2003; Siegler & Ramani, 2009; Slusser et al., 2013; Thompson & Opfer, 2008, 2010; Thompson & Siegler, 2010; White & Szűcs, 2012), although computerized versions have been used in some cases (Booth & Siegler, 2008; LeFevre et al., 2010; Newman & Berger, 1984; Pellicano, Aagten-Murphy, Attucci, Klaric, & Burr, 2011; Schneider et al., 2008; Schneider & Siegler, 2010; and Sullivan, Juhasz, Slattery, & Barth, 2011). To the best of our knowledge, touchscreen or tablet technology has been used with the number-line task in only one study (Segal, 2011), in which the focus was on the relation between gesture production and accuracy on number-line estimation. Tablets are useful not only because they are becoming more popular in classrooms but, because, in the case of number-line estimation, tabulating data collected with paper and pencil is laborious and time-consuming because it entails measuring children's marks by hand with a ruler. With tablets, data on the task can be collected accurately and instantaneously. In fact, commercially available iPad apps designed for research (e.g., EstimationLine from http://hume.ca/ix/) and for teaching (e.g., MathGlow from www.igeneration.com) have become available recently for number-line estimation. To use tablets however, we have to understand whether data obtained from tablets are likely to be similar or different from data obtained with other methods.

In this study we asked two questions aimed at illuminating whether tablets and apps are comparable to pencil-and-paper methods. First, do students perform comparably across the paper-and-pencil and tablet conditions? Second, do students prefer either presentation condition and, if so, why?

Method

To compare performance across the two presentation methods, Grade 6 students were assigned randomly to one of two conditions, tablet or paper. Students estimated the positions of numbers on a line from 0 to 100. After completing the number-line task in the assigned condition, students were asked to judge how much they liked doing the task. Next students completed five trials in the alternative condition and, finally, were asked which of the two conditions they preferred.

Participants

Thirty-two Grade 6 students (17 girls) from a suburban Canadian town were each tested individually in a 15-minute session. Their ages (in years;months) ranged from 9;11 to 13;00 with a mean of 11;10.

Number-line estimation task

The task was adapted from the number-to-position estimation procedures of Booth and Siegler (2006) and Barth and Paladino (2011). For the 0- to-100 range, two targets were selected from every decade. Stimuli in the assigned condition were presented as two sets of 10 targets, one target per decade, with a short break between sets. Targets were ordered unsystematically with the constraint that no targets from the same decade appeared in adjacent positions, and targets from adjacent decades did not appear in consecutive trials. Two presentation orders of trials were created, the second the reverse of the first, and counterbalanced with gender and presentation condition (Table D1). Following these 20 trials, students were exposed to the task in the alternative presentation condition so that they could indicate which presentation condition they preferred. In the alternative presentation condition students estimated an additional 5 targets, which was sufficient to provide a sense for the alternative format.

Table D1

Stimuli for Number-line Estimation Tasks

Presentation Mode	Set Order	Target Values: Order A[a]	Target Values: Order B[a]
Primary Mode (iPad or Paper)	Set 1	21, 48, 61, 94, 57, 36, 88, 17, 8, 76	76, 8, 17, 88, 36, 57, 94, 61, 48, 21
	Set 2	96, 33, 4, 52, 81, 42, 14, 67, 25, 72	72, 25, 67, 14, 42, 81, 52, 4, 33, 96
Opposite Mode (Paper or iPad)	Set 3	93, 78, 7, 53, 28	

[a] Targets were ordered semi-randomly with the constraint that no targets from the same decade or century appeared in adjacent positions, and that trials did not appear in ascending or descending order. The second order is the reverse of the first.

Conditions

The number-line estimation task was presented either as an app on a tablet or with paper and pencil. On an Apple iPad, number-line stimuli were presented as a straight line on the touch screen, 18 cm in length, with 0 marked at the left end and 100 at the right end. The target number was centered 2.5 cm above the line. To make their estimates on the iPad, participants used a Kuel H10 high sensitive stylus. With paper and pencil, the same configuration of the number-line and target described above were used except that the trial was presented on an 20.3 x 30.5 cm (8.5 x 11 inches) sheet of paper, with a black border around the number line to mimic the tablet display, instead of a touch screen, and children used a pencil instead of a stylus.

Procedures

Instructions were adapted from Booth and Siegler (2006) and Barth and Paladino (2011) and were identical for both conditions except for references to the screen/paper and the stylus/pencil. Students were told:

> "I want you to show me, by marking with the {pencil, stylus} on this line from 0 to 100, where you think the number at the top of the {page, screen} goes on the number line. Make your best answer as quickly as you can and tell me "done" when you have your answer. I will tell you to {circle your, press the ok button for your} final answer and go to next one."

On the tablet, we took advantage of some of the capabilities afforded by the technology. For example, when the number line appeared the first time, the endpoints were displayed in red and enlarged 200% and then scaled down to the normal size. These dynamic numbers were meant to draw attention to and orient the student to the current number-line range. In the paper condition, the experimenter oriented students to the number-line range by simply pointing to each endpoint as she spoke. In both conditions, attention was drawn to the end points of the line although the dynamic display of the tablet allowed us to graphically draw attention to the endpoints. Finally, students practiced making marks by using the stylus or pencil to show the experimenter where the practice targets, 0 and 100, were on the number line.

Students then estimated the location of 10 targets, each on a separate 0-to-100 number line. After completing the first set, students in the tablet condition saw a screen that appeared with a video of some cheerful music and animated dancing bunnies holding a sign that said “Nice Job!” In the paper condition the experimenter told students, “Nice job!” In both conditions, students were given positive feedback for completing a set and again, the dynamic display of the tablet afforded the opportunity for a short video display. Students in both conditions then estimated the second set of 10 targets.

At the end of both sets students in both conditions were shown a scale from 1 to 7 and asked by the experimenter, “I’m curious to know, on a scale of 1 to 7, where 1 is really dislike (pointed to picture of frowning face icon), 4 is neither like nor dislike (pointed to picture of neutral face icon), and 7 is really like (pointed to picture of smiling face icon), how much did you like doing number lines {on the iPad, with paper and pencil}?” After recording the student’s rating, the experimenter asked “Why?” and recorded the student’s answer.

Finally, to expose students to the alternative presentation mode, students in the tablet condition estimated five targets with paper and pencil, and students in the pencil-and-paper condition estimated five targets with the tablet. After these trials the experimenter asked “Which did you like doing number lines on more: the iPad or with paper and pencil?” After recording the student’s preference, the experimenter asked “Why?” and recorded the student’s answer. At the end of the session, the experimenter thanked the student.

Results and Discussion

Number-line estimation

To determine whether accuracy on the number-line estimation task varied as a function of estimating with pencil and paper or on a tablet, each child's proportion of absolute error (percent absolute error or PAE; in Booth & Siegler, 2006, 2008) was calculated as:

$$PAE = \frac{|\text{Estimated Position} - \text{Target Presented}|}{\text{Numerical Range}}$$

For example, if a child was asked to estimate the location of 42 on a 0-to-100 number line and placed the mark at the location that corresponded to 55, the proportion of absolute error on that trial would be calculated as (|42-55|/100), equalling .13, or multiplied by 100 to reveal a percent absolute error of 13%.

Each student's mean percent absolute error across 20 trials was calculated. Generally errors were small, with an overall mean PAE of .04. A between-subjects Gender (Male, Female) x Condition (Tablet, Paper) ANOVA was conducted to test for differences in students' accuracy when estimating on paper compared with estimating on a tablet. No main effects or interaction were found, Fs$(1,28) \leq 1.74$, ps $\geq .20$, η^2s $\leq .06$. The high level of comparability between the two conditions is illustrated in Figure 1.

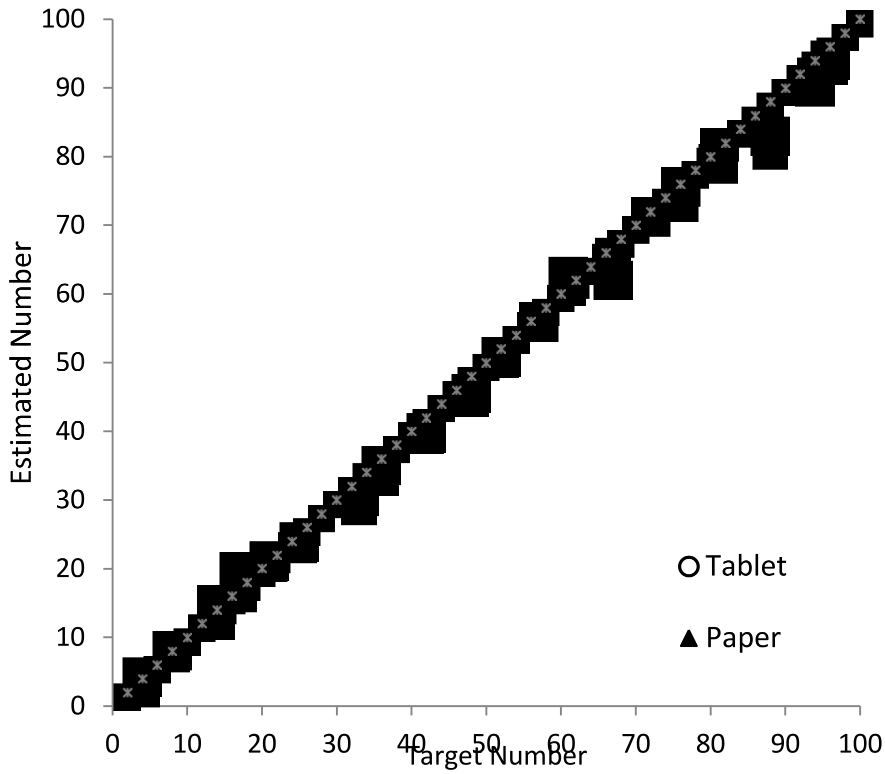


Figure D1. Median number-line estimates for Grade 6 students in one of two presentation conditions, tablet or paper. The grey crosses indicate perfectly linear performance.

In Figure D1 median estimates are plotted as a function of each target's true position, confirms the similarity in students' performance when using either paper or the tablet. Initial differences between the conditions may have been obscured by averaging across 20 trials. To test for this possibility, we calculated the mean absolute error for each block of five targets. A Condition (Tablet, Paper) x Block (1, 2, 3, and 4) ANOVA with repeated measures on the last factor revealed no effects. Moreover, a planned comparison on the mean absolute error between the first and last blocks showed no difference (M = .04, SD =.02 in both blocks). Students performed equally well throughout the task irrespective of the presentation mode.

Students' Preference

When asked how much they liked "doing the number lines," students rated both presentation conditions as equally likeable (M = 5.63, SD = .81 for tablet; M = 5.31, SD = 1.10 for paper and pencil) (p =.36). Students then estimated five targets in the alternative presentation condition and were asked whether they preferred estimating with pencil and paper, or with an app. Nearly three-quarters (23 of 32) of students preferred the app. The binomial probability of obtaining this result or an outcome more extreme by chance, is small (p = .01). A logistic regression revealed that neither presentation condition (tablet first or paper first) nor gender predicted students' preferences.

To better understand students' preferences, we considered the reasons students gave for preferring one presentation condition to the other. The explanations students' gave were sorted into features (e.g. "easy", "fun", "accurate"). Three themes emerged from examination of the features: usability, characteristics, and engagement. The frequency of these themes and features are provided in Table D2 for students who preferred the tablet and in Table D3 for students who preferred paper and pencil.

Table D2

Reasons for Preferring the Tablet

Theme (number of students who mentioned at least one feature)	Feature (number of times mentioned)	Example Responses
Usability (15)	Easier (12)	"Easier to use. Easier to make marks." "Easier to tell where to put the top number."
	Clarity (5)	"No confusion with your marks." "More simple, more clear."
	Speed (2)	"Quick and Simple." "Quicker."
	Cleaner (2)	"Sanitary." "Pencil can get messy, app was neater."
Tablet Characteristics (14)	Stylus/Tapping (8)	"I like the stylus. You just tap, so it's more simple." "Tapping is easier than drawing a line."
	Like iPad (3)	"It was fun. Like iPads." "Have one (iPad) at home, it's cooler."
	Touch Screen (2)	"I like that it's touch screen." "Used to touch screens."
	Technology (2)	"Like the technology." "More fun. All electronic."
Engagement (13)	Bunnies (8)	"Dancing bunnies are cool."
	Fun (6)	"More fun." "The app is good because it's enjoyable and I'll practice more."
	Novelty (4)	"Different. I use pencil all the time, but I don't get to use iPad much." "Different than normal. Changes it up so you enjoy it more."

Note. This information comes from the 23 of 32 students who preferred the tablet to paper and pencil.

Table D3

Reasons for Preferring Paper and Pencil

Theme (number of students who mentioned at least one feature)	Feature (number of times mentioned)	Example Responses
Usability (9)	Accuracy (8)	"Line is more accurate. Clicking the touch screen made slightly different marks than desired." "More direct. Sometimes the stylus doesn't put the mark where you want it to be."
	Control (3)	"You have more control over the mark." "Easier pencil and paper doesn't slip."
	Speed (2)	"Quicker. Make a line instead of clicking through trials."
	Easier (1)	"Easier."
Paper Characteristics (2)	Multiple marks (1)	"Can make more tic marks on paper."
	Familiarity (1)	"Familiar with paper and pencil."
	Tactile/Labeling (1)	"Paper more hands on, can label numbers."

Note. This information comes from the 9 of 32 students who preferred the paper and pencil to the tablet.

Several findings are notable. First, in both presentation conditions children listed features associated with the general theme of usability. More than half (65%) of the students who preferred the tablet mentioned features associated with usability such as speed and clarity. All nine students who preferred paper and pencil mentioned features associated with usability, particularly that they thought accuracy was better with pencil and paper. Second, more than half of the students who preferred the tablet cited characteristics such as the stylus, the touchscreen, and technology generally as part of their reasoning for preferring the tablet. In contrast, of the nine students who preferred paper and pencil, only two mentioned characteristics of the medium, such as paper being more tactile and familiar, as part of their reasons for preferring paper and pencil.

Finally, a theme unique to the tablet emerged: engagement. Of students who preferred the tablet more than half mentioned features associated with fun, novelty, and engagement. No students who preferred paper and pencil made any mention of features associated with engagement. This difference may reflect the fact that the tablet app included dancing cartoon rabbits at the end of the estimation task, as well as endpoints that were slightly more dynamic than was possible with pencil and paper. Nevertheless, the app was intentionally designed not to be particularly engaging. Tablet and computer applications more easily afford the use and integration of design features that may make the task more engaged and were part of the rationale for using a tablet for the task. Arguably, in psychological research and especially in research with children, investigators should strive to make their tasks accessible and interesting to participants. Further studies are needed to explore the relations between task engagement, task design and children's success on a task.

Conclusions

The results revealed that estimating with paper and pencil or with an app on a tablet yielded similar results, at least for Grade 6 students on a 0-to-100 number line. Grade 6 students were proficient estimators, and it may be that when faced with more difficult number-line ranges, differences in accuracy may emerge as a function of presentation mode. For familiar, well-learned ranges however, no differences in accuracy were found between the two presentation conditions.

The results also support the popular view that tablets can be engaging (Bonnington, 2012) and may be consistent with research demonstrating that students may be more willing to continue learning using mobile as compared with desktop devices (Sung & Mayer, 2013). It is not known however, whether students would be willing to engage with a task longer on tablets than on desktop computers or with pencil and paper. Children liked the tablet and paper and pencil equally but, when asked which one they preferred, most children chose the tablet. Nevertheless, tablets might not be the preferred presentation mode for all students: Almost a third of the students in our study preferred paper and pencil. Whether these results will hold for other number lines or for younger children, who may have had exposure to touchscreen technology from an earlier age, remains to be seen. Children with more exposure may find the tablet easier to use and therefore preferable but potentially less novel and less engaging. Cameron and Bush (2011), in their study of the iPad with university students, also found that the iPad was not universally preferred over pencil-and-paper methods. In future efforts to develop better psychological measurement instruments and educational interventions, it will be important to investigate how personal preference, familiarity with materials, and the affordances of new technological materials might vary systematically.

Measurement Implications

The ability to record robust data easily and instantaneously has always been a primary goal of psychological researchers. Computer tablets have considerable potential for meeting this goal. The recent success of Dufau and colleagues (2011), using a smartphone app to collect data comparable to traditional methods from more than 4,000 participants, illustrates just how powerful a measurement tool a tablet can be. Our results also confirm the comparability between data collected with a tablet and data collected with the more traditional, paper-and-pencil method.

Other information may also be collected more easily using an app such as response latencies, whether and where multiple marks are made, and the order in which marks are made. Collecting robust data on a portable, engaging, readily available computing device is exactly the kind of methodological advance psychologists should be striving to integrate into psychological research. This study serves as a base for continuing to explore the utility of tablets and apps for studying psychological phenomena. Our app and commercially available apps could be adapted to further investigate dimensions of number-line concepts such as the role of multiple representations or different kinds of representations in children's understanding and learning of the number line (Ebersbach, Luwel, & Verschaffel, 2013). For example, researchers might systematically explore the effects on children's understanding of fractions and percentages when those representations are presented in dynamic, interactive ways along the number line. Moreover, tablets allow researchers to capture different ways in which children and adults may demonstrate what they know. Being able to capture touch, auditory, video, and even eye-tracking data all on the same portable device may increase our ability to better understand the cognitive processes children use in number-line estimation.

Educational Implications

In mathematical learning especially, the use of touchscreen devices has been important because of the gestural interface they afford (Segal, 2011). Researchers have found that gesture production has the potential to enhance teaching, facilitate learning and problem solving, and act as a medium to express knowledge that children do not readily articulate (Alibali & Goldin-Meadow, 1993; Goldin-Meadow et al., 2006). An app could be designed to trace or track children's touches on the screen for the purpose of understanding how children learn concepts related to number-line estimation and how different gestures (Segal, 2011) might support learning number-line concepts.

As Dufau et al. (2011) noted, using smartphone technology "heralds a new era in behavioural sciences" with wide multidisciplinary applications. Dufau et al. listed economics, social and affective neuroscience, linguistics, and experimental philosophy. To the list may certainly be added developmental science and education. The results of this study illustrate that tablet computers and applications can contribute to measuring at least some psychological phenomena as well as traditional methods, and potentially with greater ease and in a more engaging way.

Appendix E

Coding Processes

Students' self-reports were coded for the use of twenty-one processes (see Table E1). Coders followed six steps outlined in the coding protocol at the end of this appendix. The two experimenters who administered the task coded the data by watching videos and using the coding protocol. Inter-rater reliability was 96% on a sample of data from 12 participants (four per grade).

Across both typical and atypical ranges, self-report data were collected on 2,880 trials (40 trials per student x 24 students x 3 grades). Of those, only 11 trials contained too little information to code. On another 3% of trials (15 of remaining 1,437 typical trials; 28 of remaining 1,432 atypical trials), students changed their solution processes part way through their estimations. As shown in Table E2 less than 2% of the data (25 of 1,437 typical trials; 17 of 1,432 atypical trials) were coded as "Other" responses. "Other" responses came from 15 different children. Four of these 15 children had "Other" responses on 10% or more of their trials. Two Grade 4 girls had 10% of their trials coded as "Other". One girl (ID #403) had "Other" responses over three ranges (100, 1000, and 531) and the second girl (ID #408) had "Other" responses only in unconventional ranges (220 and 531). For a third student (ID #222) 18% of his number-line estimations (across 4 ranges) were coded as "Other" because his approach to the task was generally idiosyncratic and difficult to decipher. For example, to find 325 on the number line from 0 to 1000 he said "[I] try to think of a mixture of dragons and animals…a dragon, just used wings, one wing, here and cut a little off and put the rest over here." Finally, for the fourth student (ID #419), 25% of his estimates (across 4 ranges) were coded as "Other" because he often made his estimate without fully explaining his thinking, even when prompted. For example, to find 654 on the number line from 0 to 1000 he made his mark and when prompted simply said, "54 more," and did not explain further even when prompted again. Nearly 98% of the trials from the typical ranges were coded for processes (Table E2)

Table E1

Total Number of Trials Coded by Grade

Grade	Range Type	Trials with No Information	Total Trials Coded [a]	Trials Coded as "Other"	Trials with Any Change	Trials Coded as Guess
2	Typical	3	477	11	7	13
	Atypical	7	363	7	6	29
4	Typical	0	480	12	16	6
	Atypical	1	479	10	17	22
6	Typical	0	480	1	20	1
	Atypical	0	480	0	25	3

[a] Total number of trials possible per cell is 480.

Table E2

Number of Trials Coded for Grade x Typical Range

		Grade			Total
		2	4	6	
Range	100	229	228	238	695
	1000	219	228	236	683
Total		448	456	474	1378[a]

[a] From the set of 1,412 trials, analyses of solution processes were limited to trials where students did not change their procedure during estimation (fewer than 15 trials), or guess (less 19 trials) leaving 1378 trials.

Identifying Patterns of Processes

Reformulation. Differences in types of reformulation, target or scale, were small across all ranges. Students reformulated on less than 5% of 0-100 trials and on 17% of 0-1000 trials. Trials were coded for whether are not any kind of reformulation was used as well as the specific kind of reformulation (target, scale, or both).

Anchoring Patterns. How students' used the ten specific anchors, either alone or in combination with one another, were explored to reveal 23 anchoring patterns. The task analysis of number-line estimation specifies that the anchoring process may be used once or several times to create anchoring patterns. For example, to estimate 42, one could anchor just to the middle and adjust down, or one could anchor to the 0 end, *then* to the middle, and adjust down. To establish meaningful anchoring patterns, I identified 22 anchoring patterns that were interpretable and applied those to the data. Only a few trials were unidentified and these were accounted for by creating one additional pattern, anchoring to a proportion along with another anchor. The 23 specific anchoring patterns are shown in Table E3. These 23 anchoring patterns were then collapsed into seven area-based anchoring patterns based on anchors linked to a point such as 0, the middle, or the high end, or an anchoring segment such as the whole line, or the upper or lower half of the line or to any kind of proportional anchor. Finally, the eighth general pattern accounted for use of other or constructed anchors only.

Table E3

Organizing 23 Anchoring Patterns into Eight General Anchoring Patterns

Eight General Anchoring Pattern	23 Specific Anchoring Pattern
0	
	0 only
	0 in combination with Constructed or Other anchor(s) [a]
Mid	
	Mid only
	Mid in combination with Constructed or Other anchor(s)
High End	
	High End only
	High End in combination with Constructed or Other anchor(s)
Lower half	
	Lower only
	Lower in combination with Constructed or Other anchor(s)
Upper half	
	Upper only
	Upper in combination with Constructed or Other anchor(s)
Whole line	
	Whole line only
	Whole line in combination with Constructed or Other anchor(s)
Proportion [b]	
	Proportion only
	0 in combination with at least a Proportion anchor
	Mid in combination with at least a Proportion anchor
	End in combination with at least a Proportion anchor
	Lower in combination with at least a Proportion anchor
	Upper in combination with at least a Proportion anchor
	Whole in combination with at least a Proportion anchor
	Proportion in combination with an Other anchor [c]
Other/constructed	
	Constructed only
	Other only
	Constructed & Other

[a] Constructed refers to anchors made by constructing an anchor from another anchor; Other refers to anchors that are identifiable but do not fit any other class of anchor.
[b] Proportion refers to an anchor expressed as an identifiable proportion.
[c] Identified post-hoc.

Adjustments. Ten types of adjustments were coded. The ten specific adjustment processes were collapsed into five general adjustment processes. The first two, *qualitative* and *quantitative* adjustments, were both considered "constructed" adjustments. *Qualitative* adjustments, however, were expressed more vaguely ("go up"), whereas *quantitative* adjustments were expressed in explicit, quantitative terms ("go up 2"). No differences in the frequency of use between *qualitative* and *quantitative* adjustments were found across grade so these two specific processes were collapsed into the general *constructed* adjustment process. The three specific *count by* ___ adjustments (*counting by ones*, *tens*, or *other*) were each used too infrequently to be analyzed in detail so they were collapsed into the general *count* adjustment process. *Middle* and *proportional* processes were combined into the general *proportion* adjustment process because finding the middle is simply a specific kind of adjusting proportionally. The process of adjusting by using a *previous adjustment* was rarely used (on only seven trials), and was limited to four Grade 6 students. Because *previous adjustment* was so rare and grade specific, those seven trials were eliminated from the data and not considered further. The specific process of adjusting by *guessing* remained as a general adjustment process. Finally, the fifth general adjustment process was the absence of adjusting, or *none*.

Guessing. Guessing was the absence of using any anchors or adjustment and was characteristically accompanied by the child saying "guessed" or "don't know". Lack of anchoring or adjusting remained its own process.

Protocol for Coding Processes

Coders used Table E4 for coding data along with the following six steps (details below):

Step 1:Enter Time codes for start of Trial

Step 2:Briefly Transcribe relevant verbal and gesture information that may be coded

Step 3: Record kind of evidence

Step 4: Decide whether to code concurrent or retrospective report

Step 5: Code observed behavior for the *first* procedure used or described

Step 6: Other Notes

The student's first reported set of processes was taken as the "gold standard" to be as confident as possible in making assertions about how children were estimating. There are some costs to this decision. For example, we were unable to track changes in thinking within a trial that children explained retrospectively, and although these instances were noted, they were not included in the data file and not considered further. If changes in thinking were concurrent, these were noted in the coding file. Because we are not necessarily able to account for all changes in thinking, we may lose some resolution in examining possible shifting patterns of thinking across trials over time. The goal of our coding was to capture what children were thinking with as much confidence as possible and so two additional pieces of information were coded for each trial: (a) the kind of evidence (gesture and speech) the coder relied on in making her judgment, and (b) whether the child changed his or her answer or abruptly shifted to a different process mid-estimation.

Table E4

Final Coding Protocol

	EVIDENCE (binary)
Gesture	Note whether gestures are used (see specific notes below)
Speech	This should almost always be a 1 (for typically developing students, at least)
	REFORMULATE (binary)
Target	By truncation, rounding or changing form, e.g., target is 52 but refer to as 50 or half, or target is 245 but refer to as 200
Scale [a]	By truncation, rounding or changing form, e.g., scale is 0 to 100 but refer to as 0 to 10, or scale is 0 to 531 but refer to as 0 to 500
	ANCHOR (count)
0	Reference 0 endpoint
End [b]	Reference endpoint (e.g., 100 or 531)
Middle	Reference the middle, endpoint/2, ½, halfway, including the physical middle even if say wrong number (note in WRONG MID notes).
Quarter	Reference ¼ or ¾ or split in fourths (divide into 4) “25, 50,75”; counted as anchor after *Constructed Proportion* only if refer to proportion (e.g. “3/4 here”)
Other prop	Reference other proportion, e.g., third, fifth, or tenth, or split in thirds (“33,66”); can anchor to after *Construct Prop* if say proportion (e.g., “1/10th here”)
Previous Trial	Reference a previous trial – “remember where 45 was last time”
Constructed Proportion [c]	Construct anchor by counting at least 2 proportions between 2 anchors including “mid of mid” or “half of half”
Counted [d]	Construct anchor by counting at least two sequential numbers, “10, 20, 30”
Non-prop Construct	Construct anchor constructed from increment or decrement move: computation, adjustment, measurement, or between X and Y, basically - you can tell where the anchor comes from in relation to other anchors
Other [e]	Other kinds of anchors, including vague references to an area (“in the 30s”) – write in notes
	ADJUST (binary)
Qualitative	Adjust qualitatively (“up/down”; “a bit”; “some”; “near”; “far from”; “close to”, “left”, “right”, “not”, “almost”, “about”)
Quantitative	Adjust quantitatively (“add/subtract”; “about x amount”; “x mm”; “greater or less than”, “a few ticks over”)
Count by ones	Adjust by counting (“up 1, 2, 3, 4”) two or more sequential numbers
Count by tens	Adjust by counting (“60, 70”) two or more sequential numbers
Count Other	Adjust by counting (“3, 6, 9”) two or more sequential numbers
Middle	Adjust by finding “middle”, “halfway between X and Y” or “between X and Y”
Proportion	Adjust proportionally (“quarter of the way from x”, “third of way from X”)
Previous Trial	Adjust using amount from a previous trial
Guess	Adjust by guessing (“then I guessed”, “somewhere around here”, “maybe here”, “around”) or a back-and-forth gesture
	GUESS (binary)
	No use of anchors or adjusting; explicitly report “(just) guessing” or “don’t know”
	CHANGE (binary)
Answer	Change answer, “oh, wait”, “hold on”, “not what I meant”, “actually”
Strategy [f]	Change strategy or process such as “I find ¼ no wait, here is 0, 5, 10, 15, 17 there”

[a] Must be explicit or obviously implicit. If unsure, don’t code but make a note. For example, on 0-1000, counting by 100s but saying “10, 20, 30, 40” is an implicit scale change. Use of reformulation does not automatically also imply use of an anchor as in the previous example because reformulating is taken to be a separate process. Use best judgment to determine if reformulation and anchoring happen together. [b] If student says, “closer to 0 than end”, code both anchors (0 and END). [c] This column is listed as *Count Proportion* in excel files. [d] COUNT anchors may be the only and last process but differ from a counted adjustment when they aren’t anchored to anything else and act as anchors. Don’t be afraid to code count anchors and sometimes no adjustment if the counting does follow definition of an anchor. [e] When Other is the only process coded, refers to “other” things not captured elsewhere such as “I forget” or “it’s a dinosaur and you cut it at the tail and that is the number.” [f] Record details under “Adaptation and Change” section.

Six Steps for Coding Processes

Step 1:Enter Time codes for start of Trial

(Steps 2,3, 4, and 5 are parallel and iterative processes)

Step 2:Briefly Transcribe relevant verbal and gesture information that may be coded

- Children either reported their procedure **concurrently or retrospectively**
  - Concurrent = ALL evidence reported or observed as the child is thinking and before or just as she is making her first "final" mark(s) (multiple marks seen in cases of "I'm still working on this"). If there is a procedural change –code all processes and code a "1" for strategy change even if several changes to the procedure are made.
  - Retrospective is evidence reported or observed after the child's first "final mark" and may be before or after experimenter prompting.

Step 3: Record kind of evidence

- Almost necessarily, speech should a "1"
- Gestures to consider as additional evidence:
  - Pointing (such as when anchoring)
  - Sweeping/area/circular gestures
  - Adjusting and counting gestures
- A note on **gesture evidence**:
  - ***If*** verbal evidence plus explicit or vague gestures
    then use and code gesture
  - ***If*** verbal evidence and gestures are MISMATCHED
    then code what you think is most accurate and rely more on what is seen (gesture or behavior).
  - ***If*** no verbal evidence but explicit gestures (e.g. pointing, adjusting, counting)
    then use and code gesture
  - ***If*** no verbal evidence and vague gestures (general direction or area motions)
    then do NOT use or code gesture

Step 4: Decide whether to code CON or RETRO

- ***If*** both verbal and gesture information in CON

  then CODE CON and review RETRO

  if same procedure ***end***

  if different procedures *Retro and add to notes ***end***

- ***If*** only verbal information in CON

  then CODE CON and review RETRO

  if same procedure ***end***

  if different procedures *Retro and add to notes ***end***

- ***If*** only gesture information in CON

  then review CON and review RETRO

  if same procedure in both, code CON ***end***

  if different procedures see gesture decisions, *Retro and notes, ***end***

- ***If*** no information in CON

  then CODE RETRO and note in Retro Only ***end***

- ***If*** no clear information in CON or RETRO

  then CODE as Other anchor = 1 and ***end***

Step 5: Code observed behavior for the *first* procedure used or described

- A **procedure** is made up of processes of anchoring and adjusting
  - **ANCHOR** = point that serves as base or constraint for subsequent operations
  - **ADJUSTMENT** = computation or qualitative move from an anchor where adjustment fixes a point from an anchor that is a new point and is the last process in estimation (else it's another anchor)
- If no anchor is clearly used, then the procedure used is guessing (**GUESS**)
- Generally, procedures will include at least one anchoring process and one adjustment process
- Sometimes procedures will include multiple anchoring processes before the final adjustment process
- Generally, the last move or process is an adjustment and necessarily everything before that is an anchoring process. That is, multiple anchors may be coded but only one kind of adjustment may be coded (because the adjustment is the last "move" before estimate)

- Words like "near" or "close by" associated with anchors are not coded as an adjustment unless the last "move" before the answer made has a qualitative referent and is then coded as qualitative adjustment. For example, target of 43 on 0 to 100 line: "it's not close to 100 or to 50 but it is close to 40", would be coded as anchor to end (100) and mid (50) and other (40) and qualitative adjustment for "close to 40".
- If an adjustment is coded there should be at least one anchor coded. That is, adjusting without anchoring is impossible but anchoring without adjusting is possible
- **Counting instances** of anchors:
    - A "1" represents the number of times a process was used (e.g., "1" in Anchor-Count column means the counting process was used one time but there may have been 6 anchors counted out).
    - Don't double count repetitions in speech that refer to the same process
    - Do count references to same anchor if used as part of a new procedure (e.g., anchoring to 0 to portion line into quarters and then anchoring to 0 again to count up by tens would result in a "2" in the Anchor-0 column and a 1 in Anchor-Quart and 1 in Anchor-Count columns).
    - Don't count "mistakes" in speech that are inconsistent with behavior – the intent of speech counts above actual speech.
    - In cases of counting anchors and then re-anchoring to a counted anchor, DO count this anchor separately as new anchor under constructed anchor (e.g., 0, 10, 20, 30, 40, 50, and then going back to the 10 anchor and counting 10, 11, 12, 13, would be coded as count anchors, followed by a constructed (going back to 10) and then adjust by counting by ones). In the case of *Constructed Proportion*, the anchor that is gone back to can be mid, quart or other prop if the proportion is specified otherwise it is a constructed anchor.

- **CHANGES** in strategy or answer
    - code as having happened or not
    - code when it is a clear a child changed how the problem is being solved or which answer is their "final" mark within a reasonable amount of time with NO verbal prompting but not so little time that when child is still working on problem, changes

an anchor or adjustment; the change is part of a mistake, e.g., in calculation or direction of adjustment

Step 6: Other Notes

Also note: details to clarify "other" answers, reminders of other things child said, the nature of count, other or constructed anchors, interesting observations, decision making points or rationale if code may be unclear, questions or points to clarify, anything else of interest, and whether response latency data is valid for that trial.

Appendix F

Identifying Procedures

To characterize students' procedures, eight general anchoring patterns and five adjustment types were combined to reveal 40 possible specific procedures shown in Table F1. For example, to find 52, a student might anchor to the middle, and then construct an adjustment, "up a little bit". This would be a specific procedure: *construct* [adjustment] *from middle* [anchor]. *Guessing*, the procedure that indicated neither anchor nor adjustment was used, was included as the forty-first specific procedure. The extent to which students used the 41 procedures for the two typical ranges, 0-100 and 0-1000 are shown in Table F1.

Table F1

Proportion Use of Specific Procedures on Conventional Ranges

Code	Anchoring Pattern	Adjustment	Specific Procedure	Proportion Use 100	Proportion Use 1000
0	-	-	Guessing	<0.01	0.02
1	0 end	None	Anchor to 0 end	<0.01	0.01
2***		Construct	Construct from 0 end	0.15	0.17
3*		Count	Count from 0 end	0.07	0.02
4		Proportional	Find Proportion from 0 end	<0.01	<0.01
5		Guess	Guess from 0 end	<0.01	<0.01
6*	Mid	None	Anchor to middle	0.02	0.04
7***		Construct	Construct from middle	0.19	0.18
8*		Count	Count from middle	0.03	0.02
9		Proportional	Find Proportion from middle	0.01	0.01
10		Guess	Guess from middle	0.01	0.01
11	High end	None	Anchor to high end	<0.01	0.01
12***		Construct	Construct from high end	0.15	0.17
13*		Count	Count from high end	0.03	0.01
14		Proportional	Find Proportion from high end	<0.01	<0.01
15		Guess	Guess from high end	<0.01	0.02
16	Lower	None	Anchor to lower half	<0.01	<0.01
17*		Construct	Construct within lower half	0.03	0.02
18		Count	Count within lower half	<0.01	<0.01
19		Proportional	Find Proportion within lower half	0.01	0.01
20		Guess	Guess within lower half	-	-
21	Upper	None	Anchor to upper half	0.00	<0.01
22*		Construct	Construct within upper half	0.03	0.03
23		Count	Count within upper half	<0.01	<0.01
24		Proportional	Find Proportion within upper half	<0.01	0.01
25		Guess	Guess within upper half	-	-
26	Whole line[a]	None	Anchor to whole line	0.01	<0.01
27**		Construct	Construct within whole line	0.05	0.06
28		Count	Count within whole line	<0.01	<0.01
29		Proportional	Find Proportion within whole line	<0.01	<0.01
30		Guess	Guess within whole line	0.01	<0.01
31	Proportion	None	Anchor to proportion	0.02	<0.01
32*		Construct	Construct from proportion	0.02	0.02
33		Count	Count from proportion	<0.01	<0.01
34		Proportional	Find proportion from proportion	<0.01	<0.01
35		Guess	Guess from proportion	-	-
36	Ad Hoc /Construct	None	Anchor to ad hoc or constructed anchor	0.01	0.02
37**		Construct	Construct from ad hoc or constructed anchor	0.09	0.08
38		Count	Count from ad hoc or constructed anchor	<0.01	0.01
39		Proportional	Find proportion from ad hoc or constructed anchor	<0.01	0.01
40*		Guess	Guess from ad hoc or constructed anchor	0.02	0.02

[a] Use of two ends and ends plus the middle constituted “whole line”. *** Denotes procedures used on at least 15% of conventional trials.** Denotes procedures used on at least 5% of conventional trials.* Denotes procedures used on at least 2% of conventional trials

Table F2 shows that the 24 most frequently used specific procedures accounted for 96% of estimates. Students almost never guessed and use of specific procedures involving proportions or proportional segments was rare, limited to 6% of trials. Most often, on 77% of trials, students used procedures that relied on landmark anchors including the 0 end, middle, and high end, or combinations or segments thereof such as the whole line or upper or lower half of the line. Occasionally, on 12% of trials, students used procedures that involved ad hoc (“20s” or “46 is here”) or constructed anchors. Table F2 shows that in addition to capturing the range of students’ specific procedures, in Grades 2, 4, and 6, on 0-100 and 0-1000 lines, students rely mostly on specific procedures at a basic or intermediate level.

Table F2

Use of 24 Specific Procedures Used on at Least 1% of Trials for Conventional Ranges

Level of Skill	General Procedure	24 Specific Procedures Capturing 96% of Trials	Proportion Use on Both Ranges
Immature	Guess	Guess	.01
Basic	0 End		.21
		Construct from 0 end	.16
		Count from 0 end	.04
		Anchor to 0 end only	.01
	High End		.20
		Construct from high end	.16
		Count from high end	.02
		Guess from high end	.01
		Anchor to high end only	.01
	Ad Hoc		.12
		Construct from ad hoc or constructed	.09
		anchor	.02
		Guess from ad hoc or constructed anchor	.01
		Ad hoc or constructed anchor only	.01
		Count from ad hoc or constructed anchor	
Intermediate	Middle		.25
		Construct from middle	.18
		Anchor only to middle	.03
		Count from middle	.03
		Guess from middle	.01
	Segments		.11
		Construct from whole	.05
		Construct from lower half	.02
		Construct from upper half	.03
Advanced	Proportional Segment		.03
		Proportion from middle	.01
		Proportion from lower half	.01
		Proportion from upper half	.01
	Proportion		.03
		Construct from proportion	.02
		Proportional anchor only	.01

Appendix G

Identifying Tactics

To identify number-line tactics I started with expectations about the kinds of tactics to code by initially drawing from the proposed sequence of tactics in the tactical ladder (Table 2). I constructed a revised tactical ladder of identified tactics after I noticed students used two tactics not originally proposed: (a) *ad hoc* procedures in several quintiles of the line, and (b) using *proportional segments* on at least one of three middle quintiles of the line in a way not dissimilar to the idea of judging part of a whole (Barth & Paladino, 2011; Slusser et al., 2013). Data were coded as shown in Table G1 based on the identified tactical ladder in the dissertation (Table 12). The criteria for Table G1 were developed in two stages. Based on the proposed tactical ladder, a description of the kinds of procedures students should be using in relation to targets within quintiles were written. These descriptions could not be applied without more precise specifications about which procedures may be included and if so, how often and in which quintiles, and if any procedures were excluded (shown in Table G1). Table G1 highlights the criteria used to judge each student's pattern of general-procedure use in relation to the quintiles of the line.

For example, on the 0-1000 number line, a Grade 2 student used *ends* on six targets and *wrong end* on four targets. Her tactic was classified as *Ends only* (Table G1). A Grade 6 student used *ends* appropriately in the first and last quintiles as well as for one target each in the second and fourth quintiles. For the remaining targets, he used *middle*. His pattern of procedures was classified as the *Landmark Middle* tactic.

Appendix G

Table G1

Criteria Used for Coding Tactics

Tactics	Six Tactical Levels	Use of Procedural Patterns
Proportional Landmarks	VI	**Include**: • *Ends* (2) or higher (4,6,7,8) at least once in each end quintile • *Middle* (4), *Segments* (6), or higher (7 or 8) in middle quintile • *Proportions* (8) at least once • May include one *Wrong End* (3) • May include *Ad Hoc* (5) **Exclude**: • *Guess* (1)
Proportional Segmenting	V	**Include**: • *Ends* (2) or higher (4,6,7) at least once in each end quintile • *Middle* (4), *Segments* (6), or higher (7) in middle quintile • *Proportional Segments* (7) at least once • May include one *Wrong End* (3) • May include *Ad Hoc* (5) **Exclude**: • *Proportions* (8) • *Guess* (1)
Landmark Middle & Segments	IV	**Include**: • *Ends* (2) or higher (4,6) at least once in each end quintile • *Middle (4)* or *Segments* (6) in middle quintile • May include one *Wrong End* (3) • May include *Ad Hoc* (5) • *Guess* (1) on no more than 1 trial **Exclude**: • *Proportional Segments* (7) or *Proportions* (8)
Ad Hoc	III	**Include**: • *Ends* (2) at least once in each end quintile • *Ad Hoc* (5) in each of three middle quintiles; at least 40% of trials overall (~4 conventional and ~2 unconventional per range) • *Middle* (4) **not** used in middle quintile • May include several *Wrong End* (3) • *Guess* (1) on no more than 1 trial **Exclude**: *Proportional Segments* (7) or *Proportions* (8)
Ends only	II	**Include**: • *Ends* (2) in each of 5 quintiles; at least 50% of trials total • May include several *Wrong End* (3) • *Guess* (1) on no more than 1 trial **Exclude**: • *Proportional Segments* (7) or *Proportions* (8)
Guessing	I	Guess on at least 30 % of trials overall (3 out of 10 conventional or 2 out of 5 unconventional)

Other	O	Does not fit any pattern above; note why
Deleted	X	40% of trials or more of trials missing (4 on conventional ranges; 2 or more trials on unconventional ranges)

One student was missing data on more than 40% of trials on 0-100 and was not coded. I applied the table formulaically to students and in cases of uncertainty, particularly in the case of *Others*, I reviewed the raw processes codes and video if needed. Review of raw process codes resulted in changing an *Other* tactic code in only one case (out of nine instances), from *Other* to an *Ends-only* tactic. Eight students were coded as using an *Other* tactic on at least one range (~5% of data). Seven students (four in Grade 2; three in Grade 4) used tactics characterized as *Other* and only one of those students, in Grade 4, used an *Other* tactic on both ranges. For example for 0-100, #203, was classified as *Other* because in the middle quintile, at targets 42 and 52, she used an *ad hoc* anchoring procedural pattern. Though she used *ends* appropriately at the ends, her use of the *middle* for 33 and *ad hoc* for 42 and 52 caused her to not fit any other tactical level. Review of her raw data for targets 42 and 52, confirmed that she did not use the middle appropriately (closer to ¾ from observed behaviour in video). For this reason, she is not deemed to have an appropriate understanding of middle. One student, #212, was classified as *Other* on 0-100 because she used *ad hoc* procedurals on every target. In reviewing the raw data, this student never explicitly used ends, or middle; instead using decade area anchors such as “20s”, “30s,” and “50s”. In all cases decades close to the target were use but she never explicitly used the middles or ends. For instance, to find 4, she gestured near, but not at the end; instead explaining that the value was near 3. For 81, she described thinking about the location of the 80s and 90s and that 81was “kind of close to 100”. This reference to decade areas or specific values, although appropriate in terms of use because they are near the targets, does not fit into any category so the tactical level of *Other*.