Adult estimation strategies:

An eye tracking investigation into reference point use for an atypical number line scale

by

Katherine M. Newman

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Master of Cognitive Science

in

Cognitive Science

Carleton University,
Ottawa, Canada

Abstract

The goal of the present study focused on the use of eye tracking technology to investigate the relationship between reference point use and performance on an atypical number line estimation task. University students (*N*=53; *M*=20.8 years) completed 33 number line trials, a post-task questionnaire, and a brief math assessment. Patterns of error and fixation data were analyzed. The results presented show that adult participants adjust their strategies in response to the scale used and that the use of implicit reference points benefits performance for targets located above the midpoint value. In sum, the varying performance demonstrated within an adult population using an atypical scale suggests that adults use proportional reasoning strategies to estimate the location of a number.

Acknowledgements

I would like to take this opportunity to thank those who have supported me through the duration of this project. First and foremost, I would like to thank my supervisor, Dr. Jo-Anne LeFevre, for her encouragement, expertise, and compassion.

Collaborations with my colleagues at the Math Lab and at the Institute of Cognitive Science have been invaluable. I would especially like to recognize Chang Xu, Heather Douglas, Matthew Huebner, Chunyun Ma, Jay Jennings, Audrey Brown and Maha Abu-Alhawa for always making time to discuss research efforts. I would like to give a special thanks to Chang and Matt for helping with experimental design and eye tracker programming.

I would also like to thank SR Research for providing Data Viewer software so that I could continue my analysis after leaving Ottawa.

My successes in life would not be possible without the love and support of my family and friends. The encouragement from my siblings and their partners, Terri, Mike, Jason, Hollie, Becky, and Aaron, has kept me motivated and focused on the future. Dave and Sheila, thank you for doing everything in your power to make life easier. Beth, thank you for planting the seed that I could take my academic pursuits all the way. To Adam and Hayden, thank you for surrounding me with positivity and bringing so much light into my life. The adventure is just beginning.

Lastly, I would like to dedicate this effort to my parents and most valued teachers, Gord and Roberta Newman. I will never underestimate the impact that you have had on my life. Thank you, thank you, thank you.

Table of Contents

List of Figures

List of Tables

List of Appendices

Introduction

Estimation involves the approximate calculation of a value, quantity, duration, or size. Adults rely on estimation abilities in day-to-day life to make rough judgments from available information (Levine, 1982; Reys & Bestgen, 1981). Such abilities are often used when splitting the bill at a restaurant, deciding the quickest route to take to work, or when parallel parking. Interestingly, estimation skills are not only necessary for approximate calculations, but are also used when deciding if a precise calculation is correct. In the present study, I used eye tracking technology to investigate the different strategies used in a number line estimation task that required the mapping of a numerical quantity onto a spatial location. Specifically, I examined the relationship between reference point use and accuracy when estimating a spatial location on an atypical number line scale.

The number line estimation task involves mapping the location of a symbolic representation (e.g., Arabic numeral) onto physical space, or vice versa (Petitto, 1990; Siegler & Opfer, 2003). In the typical number-to-position version of the task, a horizontal line is presented with vertical markers at each end (Siegler & Opfer, 2003). The participant is asked to estimate the location of a target number (e.g., 250) by marking its location on the line. As shown in Figure 1, the left end of the number line is labelled "0" and the right end is often labelled "100" or "1000". Such base-10 scales are often referred to as "typical" bounded number lines (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Friso-van den Bos, 2015; Peeters et al., 2016; Siegler & Opfer, 2003; Slusser & Barth, 2017; Thompson et al., 2013). The number line estimation task is commonly used to assess numerical knowledge in children (Barth & Paladino, 2011;

Berteletti et al., 2010; Booth & Siegler, 2006 & 2008; Cohen & Sarnecka, 2014; Geary, 2004; Geary et al., 2007; LeFevre et al., 2013; Opfer & Siegler, 2007; Rouder & Geary, 2014; Schneider et al., 2009; Siegler & Booth, 2004; Siegler & Opfer, 2003; Slusser et al., 2013; Sowinski et al., 2015). However, there are conflicting theories as to what the number line task measures (Barth & Paladino, 2011; Cohen & Sarnecka, 2014; Gunderson et al., 2012; Sullivan et al., 2011).

Figure 1. An example of a "typical" bounded number line scale.

Mental Representation or Proportional Reasoning?

Number line research has primarily focused on the changing pattern of performance observed in children (Ashcraft & Moore, 2012; Siegler & Opfer, 2003). When knowledge of the number system and relations between magnitudes are immature, children are more likely to overestimate the location of small numbers and underestimate the location of large numbers (Ashcraft & Moore, 2012; Dehaene, 2003; Dehaene et al., 2008; Laski & Dulaney, 2015; Sella et al., 2015; Siegler & Opfer, 2003; Simms et al., 2016). Children's error scores improve as they become knowledgeable about the number system and relations between magnitudes (Siegler & Booth, 2004).

Logarithmic performance. The pattern of reduced space between increasing magnitudes is known as logarithmic compression (Dehaene, 1997, 2003). Logarithmic patterns are observed for children in younger grades (e.g., kindergarten) for unfamiliar scales or scales that extend beyond their number knowledge (Siegler & Booth, 2004; Siegler & Opfer, 2003). Young children will overestimate the space between smaller numbers on a number line scale and underestimate the space between larger numbers on the scale. For example, using a 0-100 number line, kindergarten children will place small numbers (e.g., 20 and 30) farther apart than larger numbers (e.g., 70 and 80; Siegler & Booth, 2004).

Linear performance. A strong relationship has been established between linear patterns and accuracy (Ashcraft & Moore, 2012; Siegler & Booth, 2004). As children acquire number knowledge and the values on the scale are within their counting range, their estimates become more evenly spaced across the number line and are best fit by a linear function (Ashcraft & Moore, 2012; Ebersbach et al., 2008; Loftus et al., 2009; Siegler & Opfer, 2003; Simms et al., 2016). This gradual change toward more accurate, linear number line estimates is observed in the primary grades (Ashcraft & Moore, 2012; Newman & Berger, 1984; Petitto, 1990; Siegler & Opfer, 2003; Siegler & Booth, 2004). The developmental change in performance from logarithmically scaled estimates to linearly scaled estimates is referred to as the log-to-lin or representational shift (Siegler & Booth, 2004; Siegler & Opfer, 2003; Booth & Siegler, 2006; Opfer & Siegler, 2007).

Performance patterns and education. A strong association between education and linear performance has been established in several studies examining children's performance in the primary grades. Children's estimates are correlated with standardized math test scores and general math achievement (Ashcraft & Moore, 2012; Berteletti et al., 2010; Booth & Siegler, 2006, 2008; Friso-van den Bos et al., 2015; Link et al., 2013; Sella et al., 2015; Sella et al., 2017; Schneider et al., 2008; Siegler et al., 2009; Siegler & Booth, 2004; Stapel et al., 2015). Siegler and Booth (2004) showed that children's performance transitions from logarithmically placed estimates to increasingly linear estimates as children proceed from kindergarten to grade two and their understanding of numeracy develops (see Figure 2 for pattern). By the sixth grade, children demonstrate linear performance similar to that of adults (Siegler & Opfer, 2003).

Inversely, education that is delayed or absent is related to the logarithmic placement of number line estimates. The relationship between number system knowledge and linear performance on the number line task is supported by research examining the estimates of children with mathematical learning difficulties, such as developmental dyscalculia (van't Voordende et al., 2016). The inability to comprehend the number system is a core deficit of mathematical learning difficulties (Butterworth, 2010). As a result, children with mathematical learning difficulties are unable to accurately place estimates on the number line and perform below the level of their peers (Geary et al., 2008; van Viersen et al., 2013; van't Voordende et al., 2016). As well, a lack of education is related to logarithmic patterns into adulthood. Dehaene and colleagues (2008) tested Amazonian adults and children with little schooling and found that both groups showed logarithmic patterns of performance on the number line task. In sum, number line estimates are placed more accurately as number system knowledge develops.

There is little debate that children's changing performance on the number line task is related to academic achievement. However, the nature of what the number line task measures has served as the basis for competing theoretical frameworks.

The mental number line hypothesis. One theoretical position posits that changes in performance can be interpreted as the development of a linear mental representation of magnitude – a "mental number line" – and that the number line task can be used to measure the underlying mental representation of number (Ashcraft & Moore, 2012; Dehaene, 1997, pp. 83-5; Siegler & Opfer, 2003). Siegler and Opfer (2003) suggest that children rely on different mental representations of magnitude throughout development and apply different representations depending on the scale presented. For familiar number line scales (e.g., 0-10), the performance of young children is best fit by a linear model showing a fixed ratio between spatial distance and the distance between scalar values (i.e., y=x; LeFevre et al., 2013; Stapel et al., 2015). Young children's performance on less familiar scales (e.g., 0-1000) is best fit by a logarithmic function that reflects the overestimation of small numbers and underestimation of large numbers (Siegler & Opfer, 2003).

Although the mental number line theory is widely discussed in the literature, support for this view is inconclusive (LeFevre et al., 2013). A mental number line metaphor is a helpful tool for illustrating numerical concepts to children, but proponents of the mental representation hypothesis have yet to explain *how* cognitive processes operate on such representations (Ashcraft & Moore, 2012; Dehaene, 1997; Siegler & Opfer, 2003).

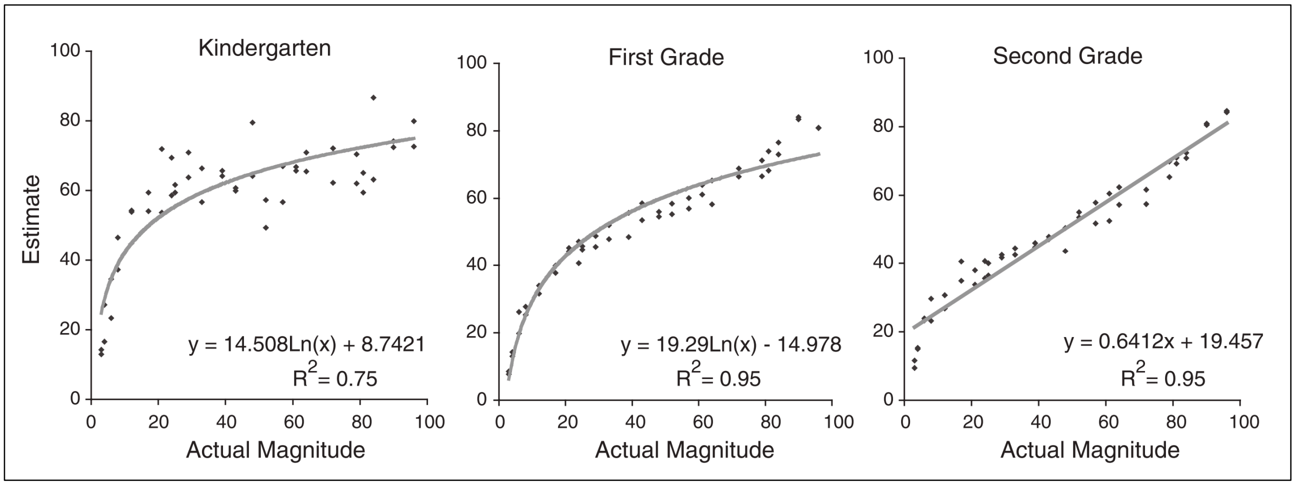

Figure 2. Results from Siegler & Booth (2004) showing the progression from logarithmic (kindergarten) to linear pattern (second grade).

Note. Median estimates used.

The proportional judgment account. An alternative position, suggested by Barth and Paladino (2011), is that this change in performance results from an increase in number knowledge and proportional estimation abilities involving the use of explicit and implicit reference points (Huber et al., 2014; Link et al., 2014; Slusser & Barth, 2017; Stapel et al., 2015; Sullivan et al., 2011; van't Voordende et al., 2016).

The visual-proportional nature of the number line task encourages the use of spatial reasoning processes specific to visual processing (Landy et al., 2013). Estimating locations on the bounded number line task requires proportional reasoning ability for participants to make relative judgments. A target location cannot be determined by estimating numerical magnitude alone. The location of the target number must be judged in relation to the value and position of the high endpoint (e.g., 50/100; Barth & Paladino, 2011; Slusser & Barth, 2017; Peeters et al., 2016). The accurate placement of a target number can be further improved through the use of self-generated, implicit reference points (e.g., the midpoint; Peeters et al., 2016). Schneider et al. (2008) suggest that the ease associated with line bisection strategies may predispose participants to use a midpoint reference point. Using a midpoint strategy is especially effective; several studies have shown that a midpoint referent is associated with reduced error for targets located near this location in comparison to other targets (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Mock et al., 2016; Schneider et al., 2008; Siegler & Opfer, 2003; Siegler & Thompson, 2014; Sullivan et al., 2011). As accuracy increases with age and experience, patterns of error signal that both explicitly marked (e.g., endpoints) and implicit (e.g., midpoint) reference points are used to serve as anchors from which estimates are adjusted (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Peeters et al., 2016; Schneider et al., 2008; Siegler & Opfer, 2003; Sullivan et al., 2011). Therefore, people are expected to show the greatest amount of error on targets located the farthest from commonly used reference points (e.g., the endpoints and midpoint; Ashcraft & Moore, 2012; Link et al., 2014; Petitto, 1990; Rouder & Geary, 2014; Siegler & Opfer, 2003; Slusser & Barth, 2017). This pattern was noted by Ashcraft and Moore (2012) who observed an "M-shaped" pattern of reduced error in the endpoint and midpoint regions when error scores were graphed. Sullivan et al. (2011) found support that endpoint and midpoint reference point use was related to a decrease in error when fixation data was graphed, producing a complimentary, "W-shaped" pattern of increased fixations in these regions (see Figure 16A). In Barth and Paladino's (2011) view, shifting patterns of performance need not appeal to mental representations when education and the proportional structure of the bounded number line are taken into account.

Proportional judgment on an atypical number line scale. Given that previous research has found that the scale of the number line likely influences the strategic reference points used by adults, an “atypical”, non-base-10 scale was chosen for this study to investigate whether reference point use could be manipulated (Landy et al., 2013; Siegler & Opfer, 2003). In the present research, a bounded number line scale ranging from “0” on the left to “7000” on the right was used for all trials. Considering that adults are familiar with the values included in a 0-7000 scale, it was expected that they would be able to make use of relevant reference points and linearly place their estimates. However, the manipulation of the scale was intended to challenge the proportional judgment skills of participants. For example, it was expected that estimating the location of 5600 on a 0-7000 scale would be more difficult than estimating the location of 800 on a 0-1000 scale, even though both targets are located at 80% of their respective scales. I hypothesized that an atypical number line scale would increase variability not only in performance, but also in the kinds of reference points used. I anticipated that, in addition to using the endpoints and the midpoint, some participants would be more likely to use thousand markers as reference points on a 0-7000 scale. Based on previous findings that reference point use is related to less error, I expected error scores to be lower for participants who use these reference points (Barth & Paladino, 2011; Newman et al., 2017; Sullivan et al., 2011). However, if participants anchor at a reference point and adjust their estimate from left to right, then it would become more challenging to accurately make use of thousands reference points as the target values increase. As a result, mean error scores should increase for targets above the midpoint compared to targets below the midpoint and this difference would be related to the kind of reference point strategy used.

Preliminary Study

We conducted a preliminary study with 142 students (*M* age=22 years; range 18-60 years) to collect information on the self-reported strategies used in the number line task (Newman et al., 2017). Participants were tested as part of a larger study examining the relationships between math abilities, spatial skills, reading abilities, and executive functioning (i.e., inhibition and working memory). Participants completed a number-to-position version of the symbolic number line task using an iPad application (*EstimationLine*). The horizontal, bounded number line scale ranged from "0" on the left to "7000" on the right (see Figure 3).

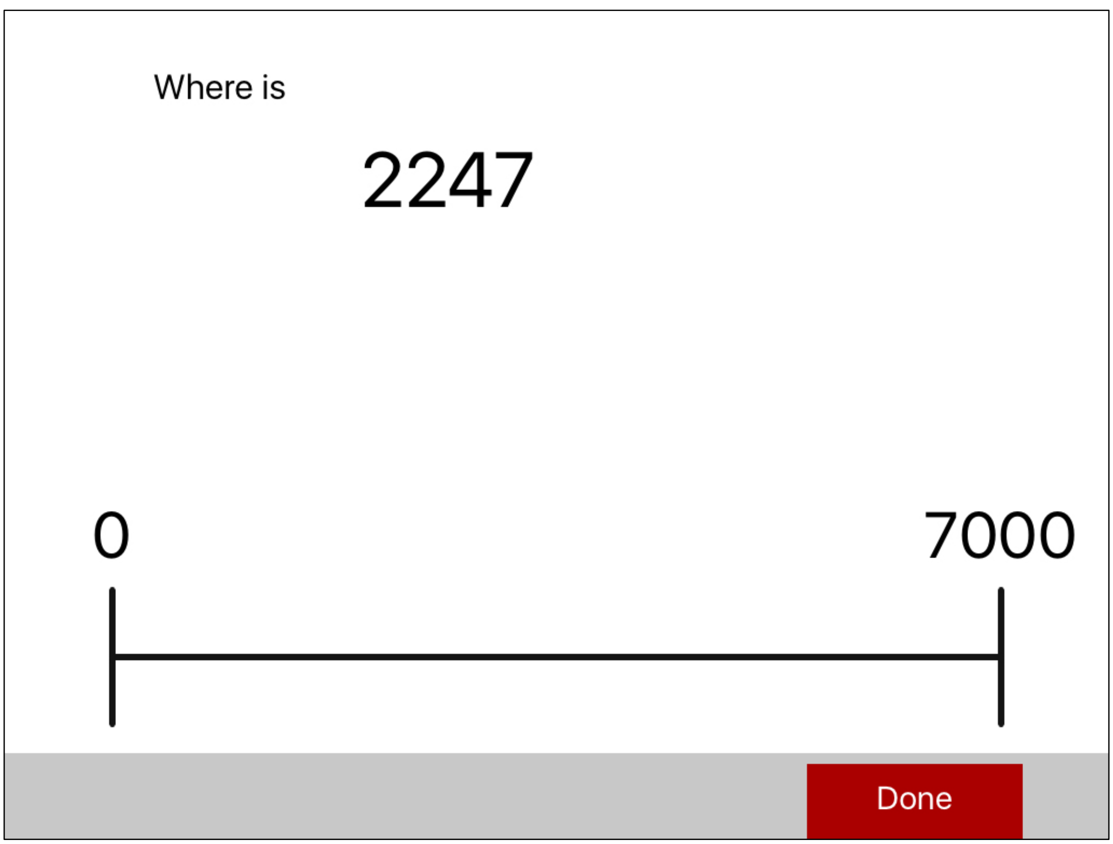

Figure 3. The "atypical" bounded number line used in Newman et al. (2017).

When the task was completed, participants were asked to report any strategies used while estimating the spatial locations. To my knowledge, this is the first study asking participants about strategy use; the majority of reference point research is inferred from behavioural data. Participants were classified as either simple strategy users (*n*=69) or complex strategy users (*n*=63) based on the number of implicit reference points reported. Simple strategy users reported using a single reference point strategy to divide the problem space (e.g., midpoint or set of spatial equivalent reference points). Complex strategy users reported using the midpoint in combination with an additional set of spatially equivalent reference points (e.g., the midpoint and thousands). The strategies reported in Newman et al. (2017) were used to inform experimental design and methods of analysis for the eye tracker version of the number line task used in the present study.

As expected, adult participants in the preliminary study were highly linear when placing their estimates. Using the target position as the independent variable and participants' median estimates as the dependent variable, curve estimation was used to fit the preliminary data (Opfer, 2003). As shown in Figure 4, a linear function fit the data (R^2=99.6%; Newman et al., 2017). The data from the present study will be similarly evaluated for model fit.

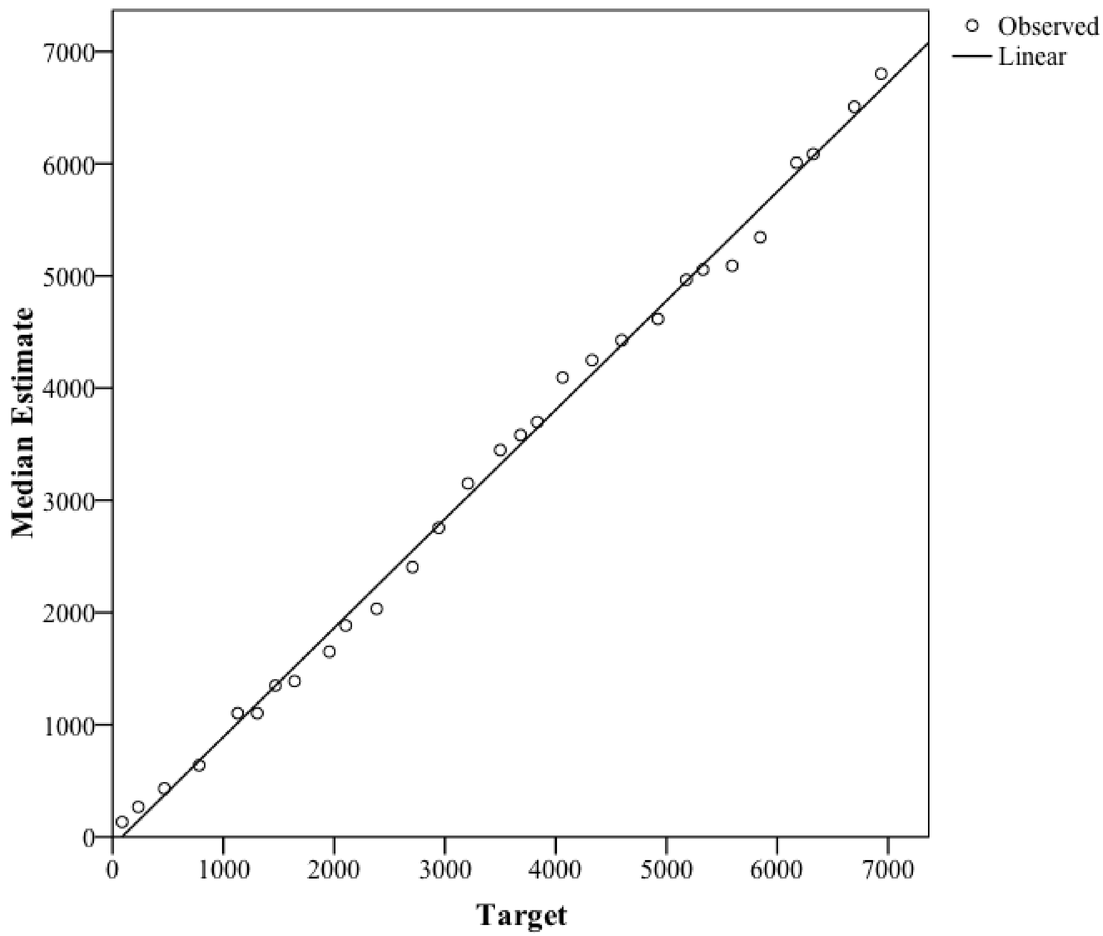

Figure 4. Linear model fit for estimates in preliminary study (Newman et al., 2017).
Note. Median estimates used.

In the preliminary analysis, a striking pattern of increasing error was observed as target values increased across the entire 0-7000 scale and within the 1000-unit intervals (Newman et al., 2017). As shown in Figure 5, mean error scores were similar for strategy groups from 0-2000 at which point the simple strategy group showed an increase in mean error as the target values increased. The complex strategy group showed significantly less error for the midpoint target, but estimates for both groups continued to increase for targets greater than the midpoint until around 6000, when maximum target values could be placed near the explicitly marked high endpoint. The M-shaped pattern observed by Ashcraft and Moore (2012) was not found for simple strategy users on the atypical number line scale. The distinct patterns of error observed for targets below and above the midpoint suggested that participants complete the atypical version of the number line task differently than the typical version using a base-10 scale. The differences in performance observed between strategy groups suggested that the use of implicit reference points was related to lower error scores, especially as target values increased in magnitude.

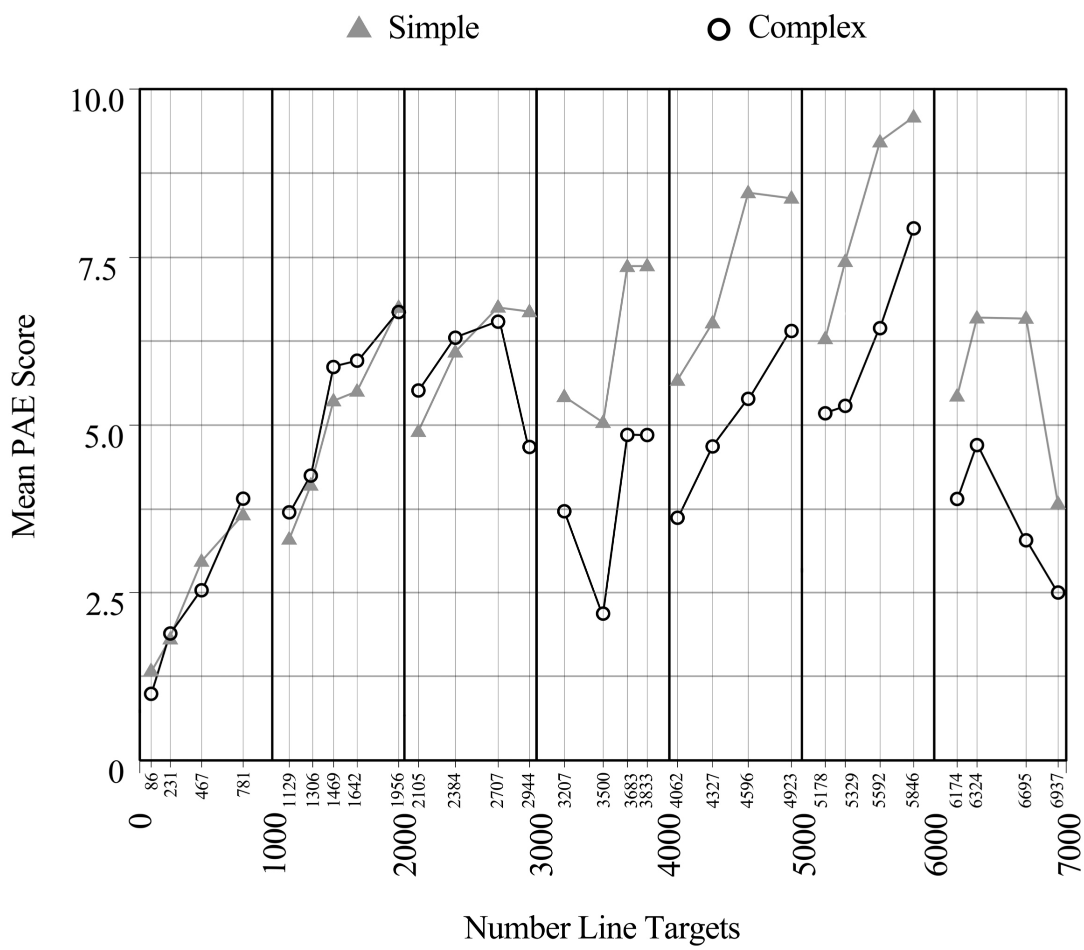

Figure 5. Mean percent of absolute error scores for number line targets separated by strategy group (Newman et al., 2017).

To investigate the relationship between target location and error, we examined mean error scores in two ways: i) for targets below and above the midpoint of the entire scale (e.g., targets < 3500 compared to targets > 3500), and ii) for targets below and above the midpoint of each interval of 1000 units (e.g., targets between 1000-1500 compared to targets between 1500-2000; Newman et al., 2017). Results indicated that error scores were higher for targets above the midpoint compared to targets below the midpoint (e.g., 5846 had a higher mean error score than 1956). Error scores were also higher for targets above the midpoint of each thousand interval (e.g., 4923 had a higher mean error score than 4062). When error scores were analyzed by strategy group, it was revealed that complex strategy users had lower mean error scores overall and that their error scores were significantly lower than simple strategy users for targets above the midpoint. Complex strategy users also had less error for targets above the midpoint of each 1000-unit interval. The complex strategy group was significantly more accurate when locating the midpoint target compared to the simple strategy group. This may have contributed to lower mean error scores for targets in the upper half of the scale as complex strategy users anchored and adjusted from a more accurately placed midpoint reference. Despite the significant difference in mean error scores between strategy groups for targets in the lower and upper halves of each interval, the same pattern of increased error can be observed; error increases for both strategy groups as the value within the thousand interval increases.

These findings suggest that using more implicit reference points is related to number line estimation accuracy, especially for higher targets relative to lower targets whether within the entire scale (e.g., < 3500 vs. > 3500) or within equally sized segments of the scale (e.g., 1000-1500 vs. 1500-2000). The error patterns shared by both groups suggests that participants anchor their estimate at the thousand markers and adjust upward, from left-to-right, as the target increases within the interval. This would account for the error score increasing as the target number increases. When predictors of individual differences were considered, self-reported strategy accounted for 4% of unique variance in error on the atypical number line task after controlling for mathematical problem solving and spatial skills.

Present Study

Eye tracking methods have the potential to be especially informative when investigating the highly linear performance of adult participants on a number line estimation task (Sullivan et al., 2011). Fixation and dwell data can provide insight into individual differences in estimation abilities. The goal of the present study was to extend the results of Newman et al. (2017) to investigate eye movements using an atypical, 0-7000 scale. An equal number of target values were selected from the lower and upper halves of the scale to examine error patterns across the entire range as well as the lower and upper halves of the scale and within the 1000-unit intervals. One goal of the present study was to see how eye tracking data on the atypical number line compared to the results of Sullivan et al. (2011) on a typical (0-1000) number line. Three key findings from Sullivan et al. (2011) are of interest: i) the accuracy of the first fixation on the number line was related to estimate accuracy, ii) participants devoted the majority of total processing resources to the correct target region; and iii) participants fixate on the endpoints and midpoint regions to calibrate their estimates (see Figure 19A; Sullivan et al., 2011). The hypotheses tested in the present study concerned six key areas of investigation: multi-digit number processing, the relationship between eye movements and target regions, the relationship between first fixation and estimate accuracy, classification of self-reported strategies, reference point fixations, and predictors of individual differences in performance. A summary of my hypotheses can be found in Table 1.

Multi-digit number processing. Sullivan et al. (2011) postulated that the first fixation was influenced by the first digit in the target number and that early processing measures were predictive of later performance. However, "early processing measures" are not defined by Sullivan and colleagues (2011). In the present study, I analyzed looking behaviour for target digit regions to determine whether fixation and dwell patterns could provide insight into what early processing measures may be related to the accuracy of the first fixation on the number line. For example, does the accuracy of the first fixation on the number line relate to the target digit regions visited? Do participants who look back and forth from the target to the number line place estimates more or less accurately compared to those who do not revisit the target digit?

Empirical data related to the processing of multi-digit numbers is scant for numbers larger than two digits and is predominantly based on magnitude comparison tasks (Meyerhoff et al., 2012). The exploratory nature of the multi-digit number processing analysis in the current study also attempted to provide insight into the many competing theories of multi-digit number processing (e.g., holistic, sequential, parallel, hybrid, "chunking", etc.; see Dehaene, 1989, Poltrock & Schwartz, 1984, Nuerk & Willmes, 2005, Nuerk et al., 2001, Meyerhoff et al., 2012, and Nuerk et al., 2011 for a discussion). To investigate looking behaviour for the target number, regions of interest were defined for each of the four target digits. Based on Sullivan et al.'s (2011) suggestion that first fixations were calibrated by the first number in the sequence, I expected that participants would attend to the thousands digit first and that the majority of attentional resources would be allocated to the thousands and hundreds digits since these place values were the most relevant when estimating the location of a target. I also investigated whether participants looked back-and-forth between the target digits and the number line.

Relations of fixations and dwell times to target number. Fixation and dwell data were also analyzed for three regions of interest located around each target number to determine whether eye movements were related to estimation accuracy: underestimated, correct, and overestimated. Sullivan et al. (2011) found that participants devoted approximately half of their processing resources to these three regions, comprising 15% of the number line, around the target location (see Figure 11). Although previous studies have analyzed fixation data for larger target regions, Sullivan and colleagues (2011) point out that these studies were concerned with children's performance and that the large, previously defined regions of interest are not appropriate for highly accurate adult participants. The more precise target regions of interest used by Sullivan et al. (2011) and in the present study are better suited to analyze the subtle fixation patterns demonstrated by adults.

A variation between the following analysis and that conducted by Sullivan et al. (2011) is that conditions were not used in the present study. Sullivan et al. (2011) separated participants into small- and large-initial-number conditions. Participants in the small-initial-number condition completed all trials smaller than the midpoint (e.g., < 500) first, followed by targets larger than the midpoint (e.g., > 500). These conditions were used to determine whether participants estimate the target location based on previous estimates, or if estimates are calibrated online. If participants rely on memorized correspondences, then initial-number condition should not bias their estimates. If estimates are calibrated online, then the initial-number condition may bias participants' estimates. Sullivan et al. (2011) found that participants in the small-initial-number group devoted most of their processing resources to the correct region around the target number and had more fixations in the overestimated region compared to the underestimated region. Participants in the large-initial-number condition also allocated the majority of processing resources to the correct region, but fixation and dwell data did not significantly differ between the correct region and underestimated region. In sum, those in the small-initial-number condition were more likely to look in the overestimated region compared to the underestimated region while those in the large-initial-number condition were more likely to show the reverse pattern. The dissimilar patterns of looking behaviour for small- and large-initial-number conditions lead Sullivan et al. (2011) to conclude that adults calibrate their estimates online.

A limitation worth noting is that Sullivan et al. (2011) included only eight participants in each condition. The small sample size used calls into question whether these results can be replicated. I opted not to separate participants into conditions, but to examine if the relationship between eye movements and the three target regions differed for the lower and upper halves of the scale. I expected eye movement data to be consistent with Sullivan et al.'s (2011) finding that the majority of adult participants' eye movements were allocated to the correct target region. However, I also hypothesized that participants would devote more processing resources to the underestimated region for targets above the midpoint in comparison to targets below the midpoint. The pattern of underestimating targets above the midpoint was also observed by Newman et al. (2017).

Relations between first fixation and estimate accuracy. Fixation data was analyzed in an effort to replicate another of Sullivan et al.'s (2011) findings: the relationship between first fixation and estimate. The first fixation on the number line was found to predict estimate accuracy (Sullivan et al., 2011). In the present analysis, I examined the relationship between first fixation and estimate accuracy for targets below and above the midpoint. Considering that participants were more likely to underestimate their estimates for targets above the midpoint in the preliminary study, I hypothesized that the relationship between first fixation and estimate accuracy would depend on whether the target was located in the lower or upper half of the scale (Newman et al., 2017). Specifically, I expected that participants' first fixations would be less accurate for targets above the midpoint and that this would contribute to the likelihood of a higher error score for the final fixation.

Classification of self-reported strategies. Self-reported strategies were collected following the number line task and classified as simple or complex based on the number of implicit reference points reported (e.g., midpoint strategy vs. midpoint and thousands strategy). Strategies within classification groups were analyzed to ensure that performance did not significantly differ between them (e.g., within the complex group, participants who reported using the midpoint and thousands did not significantly differ from participants who reported using the midpoint and quarters). Mean accuracy was examined by strategy group for the entire scale (e.g., 0-7000), for targets below and above the midpoint (e.g., < 3500 compared to > 3500), and for low- and high-thousands targets (e.g., 1000-1500, 2000-2500, 4000-4500, 5000-5500 compared to 1500-2000, 2500-3000, 4500-5000, 5500-6000; intervals that included the endpoints and midpoint were not analyzed). Consistent with results from the preliminary study, I expected that complex strategy users would have lower mean error scores overall, for the lower and upper halves of the number line scale, and for the low- and high-thousand targets compared to simple strategy users (Newman et al., 2017). This result was expected because complex strategy users report using more implicit reference points to help divide and further subdivide the spatial area so that estimates can be located more accurately. Differences in performance were not anticipated for specific strategy types within the simple and complex groups.

Reference point fixations. The analysis of reference point fixations in the present study was conducted to establish which reference points are most commonly used on an atypical, 0-7000, number line scale. Although a number of studies have shown a decrease in error for targets located near the endpoints and midpoint, few have examined fixation data for reference point regions used by adults (Ashcraft & Moore, 2012; Heine et al., 2010; Schneider et al., 2008; Sullivan et al., 2011). To examine reference point use beyond the self-reported strategies, fixation data was analyzed for nine commonly reported reference points: 0, 1000, 2000, 3000, 3500, 4000, 5000, 6000, and 7000. I hypothesized, due to the atypical scale used, that participants would fixate on the thousand markers in addition to the midpoint and endpoint locations. Complimentary to the self-report data, I expected that complex strategy users would have a higher number of overall fixations as well as fixations in more reference point regions compared to simple strategy users.

Predictors of error for lower and upper regions of the number line scale. The final analysis in the present study aimed to identify factors that account for individual differences in error on an atypical number line task. Due to the variation in performance observed for targets below the midpoint and targets above the midpoint by Newman et al. (2017), these regions were analyzed separately. Fraction ability, strategy group, and reference point fixation data were considered predictors. Based on the results from Newman et al. (2017), I anticipated that fraction ability and strategy group would predict performance for targets both below and above the midpoint. Because complex strategy users report using more reference points, I expected this group to have more fixations in reference point regions above the midpoint and that reference point use would predict a decrease in error for targets within this more challenging region.

Table 1

List of hypotheses by areas of investigation

Area of Investigation	Hypothesis
Multi-digit number processing	1. Participants will first fixate on the thousand digit of the target number
	2. Most target number fixations will be in the thousands and hundreds digit regions
Relations of fixations and dwell times to target number	3. Most number line fixations will be in the correct target region
	4. Participants will look more in the underestimated region for targets above the midpoint
Relations between first fixation and estimate accuracy	5. The relationship between first fixation and estimate accuracy will depend on target location (e.g., below or above the midpoint)
	6. First fixations will be less accurate for targets above the midpoint
Classification of self-reported strategies	7. Error will not significantly differ among strategies belonging to the same strategy group
	8. Complex strategy users will have lower mean error scores for targets above the midpoint compared to simple strategy users
	9. Complex strategy users will have lower mean error scores for targets above the interval midpoint compared to simple strategy users

Reference point fixations	10. Complex strategy users will have more overall fixations and more fixations in reference point regions compared to simple strategy users
	11. Participants will be more likely to use reference points located at the thousand markers for an atypical number line scale
Predictors of error for lower and upper regions of the scale	12. Fraction ability and self-reported strategy will predict mean error for targets on both the lower and upper halves of the scale
	13. Reference point fixation data will predict error for targets above the midpoint

Method

Participants

Fifty-three students from Carleton University (29 males) participated in the study for course credit. The mean age was 20.8 years (*SD* = 3.5 years; range 18-35 years). All participants reported normal or corrected-to-normal vision. Eighty-three percent of participants spoke English as their first language. Other first languages reported were French (3.8%), Mandarin (3.8%), Arabic (3.8%), Japanese (1.9%), Vietnamese (1.9%), and Karen (1.9%). Data of three participants were not included in the reported analyses. One participant was not included because she completed 31/33 trials believing that the scale ranged from 0-10,000. Fixation data for this participant indicated that she did not fixate in the upper quarter of the number line scale. Additionally, two participants who did not report using a reference point as part of their strategy were not included in any analyses. Thus, the data for 50 participants was available for further analyses.

Materials

Mathematics Beliefs and Interests Questionnaire (MBIQ). Testing began with the participants completing the MBIQ. The MBIQ is used to gather demographic information (i.e., primary language, sex, handedness) as well as perceived confidence ratings of math, spatial, and reading abilities. The questionnaire was hosted on Survey Monkey and data was collected anonymously.

Bounded number line task. Participants completed 33 trials on an atypical (0-7000) bounded number line estimation task. Stimuli were presented on a 15-inch monitor with a resolution of 1024 x 768 pixels. Target numbers and the number line were presented in black on a white background (see Figure 1). The number line scale ranged from 0 on the left to 7000 on the right. The 7000-unit line had a physical length of 750 pixels (38.5 cm). As shown in Figure 6, target numbers were displayed in the top left corner of the monitor.

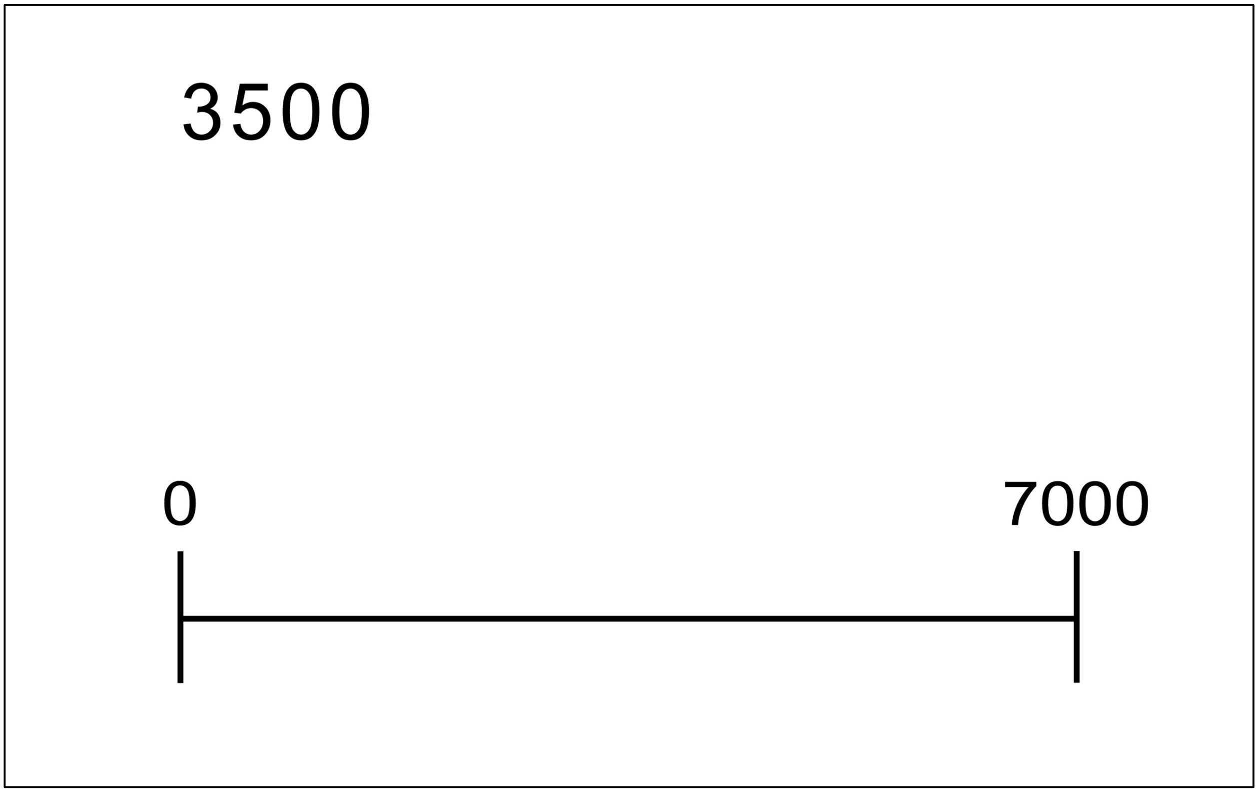

Figure 6. Target number and number line stimuli presented on computer monitor.

Thirty-three target numbers were selected to cover the entire number line scale as well as the range within each thousand-unit interval. The targets were divided into two balanced sets. As shown in Table 1, Set 1 consisted of 16 targets sampled equally from below and above the midpoint value. Set 2 consisted of 16 targets equally sampled from the lower and upper halves of the number line scale. Half of the participants saw Set 1 first and Set 2 second; the other participants saw the two sets in the reverse order. The midpoint value (3500) was presented between Sets 1 and 2 for all participants. The order in which items were presented within each block of trials was determined randomly for each participant.

Table 2

Number line Stimuli

	Thousands Unit						
Set	0000s	1000s	2000s	3000s	4000s	5000s	6000s
1	86	1642	2105	3402	4062	5178	6174
1	467	1956	2707	3683	4596	5261	6695
1					4782	5592	
				3500			
2	231	1129	2384	3207	4327	5329	6324
2	781	1306	2944	3609	4923	5846	6937
2		1739		3833			

Note. Order of sets was counterbalanced across individuals. All participants saw the target 3500 between the two problem sets.

Post-task questionnaire. Immediately following the completion of all number line trials, participants were asked to report any strategies used to locate their estimates (see Appendix A for full questionnaire). Participants who reported using the midpoint as a reference point were asked to report the value of the midpoint. Participants who did not report using the midpoint were asked to report the value of the midpoint at the end of the interview. Participants were then asked to answer questions about their solution strategies for three specific targets: 2707, 5178, and 6695. Questions for each target pertained to reference point use and number rounding. All responses to the post-task questionnaire were recorded with the participant's permission and later transcribed by the experimenter.

Math ability. Participants' mathematical ability was assessed using an extended Brief Math Assessment (BMA-3; Appendix B) that included whole-number, fraction, and algebra arithmetic problems. Three new items were added to the original BMA-3, as described below. The original 10-item BMA-3 was developed by Steiner and Ashcraft (2012) and was based on the Wide Range Achievement Test: Third Edition (WRAT3). They reported a reliability of .69 (Cronbach's alpha); for the same 10 items, the reliability in the present sample was .63. Three additional items that involved fractions were added to the set of questions used by Newman et al. (2017) because the fraction items showed strong relations with number line performance. The BMA-3 Revised contains 13 math problems (Cronbach's α=.59) ranging in level of difficulty. As shown in Appendix B, items 8, 10, and 12 were added. The second equation in item 11 was also updated to make it more challenging for participants. Problems included whole number addition, subtraction, and multiplication (n = 4), addition and subtraction of mixed fractions (n = 4), decimal to fraction conversion (n = 2), and algebra (n = 3). Participants were given a maximum of ten minutes to complete all items. Scoring is based on the total number of correct responses. A subscore based on the six fraction and decimal-fraction conversion questions was also calculated; for these, reliability was .56.

Apparatus

Number line stimuli were visually presented on a 15-inch Dell monitor attached to a Dell Precision PWS390 computer, using Windows XP Professional 2002 operating system to run EyeLink 1000 experiment software (SR-Research, Kanata, Ontario, Canada). The computer monitor logo and stand were covered to avoid cueing the participant to the middle of the screen. A specialized camera was used to record participants' eye fixations and saccades while completing the number line task. The camera emitted a beam of infrared light that was reflected off the participants' cornea at a sampling rate of 1000 Hz.

Procedure

Participants were assessed individually on all measures. After reading and signing the informed consent document, participants completed the MBIQ. Testing began with the number line estimation task. Participants were seated approximately 68 cm from the infrared camera and 110 cm from the monitor (see Figure 7 for set up). The chair height and chin-rest were adjusted for the comfort of each participant and to ensure that the camera had a full view of the participants' eyes. Before the experiment began, instructions were presented on the monitor as the experimenter read them aloud. A nine-point calibration was completed followed by two practice trials. When the practice trials were successfully completed, the calibration procedure was repeated and the participant was instructed to maintain a still position until experimental trials were completed. For the duration of testing, participants kept the index finger of their dominant hand on the "ENTER" key. To begin each trial, participants were instructed to press "ENTER". A gaze-contingent black square would then appear in the top right corner of the screen. Once a fixation was recorded in the square, the square would disappear and the target number and number line would be presented simultaneously. All target numbers were presented in the same location in the top left corner of the screen (see Figure 6).

Participants were instructed to look at the target number and then "quickly and accurately" estimate the target's spatial location on the line. When the participant decided on the target's location, they would fixate on that location and press "ENTER". This process would be repeated for subsequent trials.

Figure 7. Eye tracker setup showing location of monitor, camera, keyboard, and chin-rest in relation to participant.

The number line post-task questionnaire was conducted immediately following the completion of all trials. Participants were instructed to remove their head from the chin-rest and relax before the experimenter asked them a few questions about how they completed the task. Participants were informed that the interview would be recorded.

The participant was first asked to report any strategies used at the task-level (see Appendix A for full questionnaire). Participants were also provided with three targets from the experiment and asked to report any trial-level strategies used. The three targets were presented in the same order for each participant: 6695, 5178, and 2707. For each target, the participant was shown a piece of paper with the target number displayed in Arial font, size 200, while the experimenter asked a series of questions pertaining to the target. When the interview was completed the experimenter stopped the recording.

The BMA-3R was completed in a separate testing room. Testing commenced when all questions were completed, or when 10 minutes had elapsed. After completing the experiment, participants were debriefed and dismissed.

Data Analysis

For the 50 participants included in the analyses, 118/1650 (7.15%) trials were missing fixation data (i.e., the participant pressed ENTER before fixating on the number line). These trials were removed from analyses.

Results

For all analyses of variance (ANOVAs) where there were more than two levels of an independent variable, Greenhouse-Geisser corrections were used if sphericity was violated and Bonferroni corrected *p* values were used when comparing means for significant effects.

On average, participants made 6.3 fixations on the number line per trial and the average fixation duration was 490 ms. Accuracy of the estimate was calculated as percentage of absolute error (PAE) using the formula: [(Participant's estimate – Target location)/Scale of estimate] x 100 (Petitto, 1990). For example, if a participant was presented with a target of 500, but their estimate corresponded to a location with a value of 1000, the PAE score was 7.14% (i.e., [(1000 – 500)/7000] x 100). The reliability analysis, based on PAE for all 33 number line trials, showed a Cronbach's α of .79. Thus, performance on this task was similar to that shown by Newman et al. (2017).

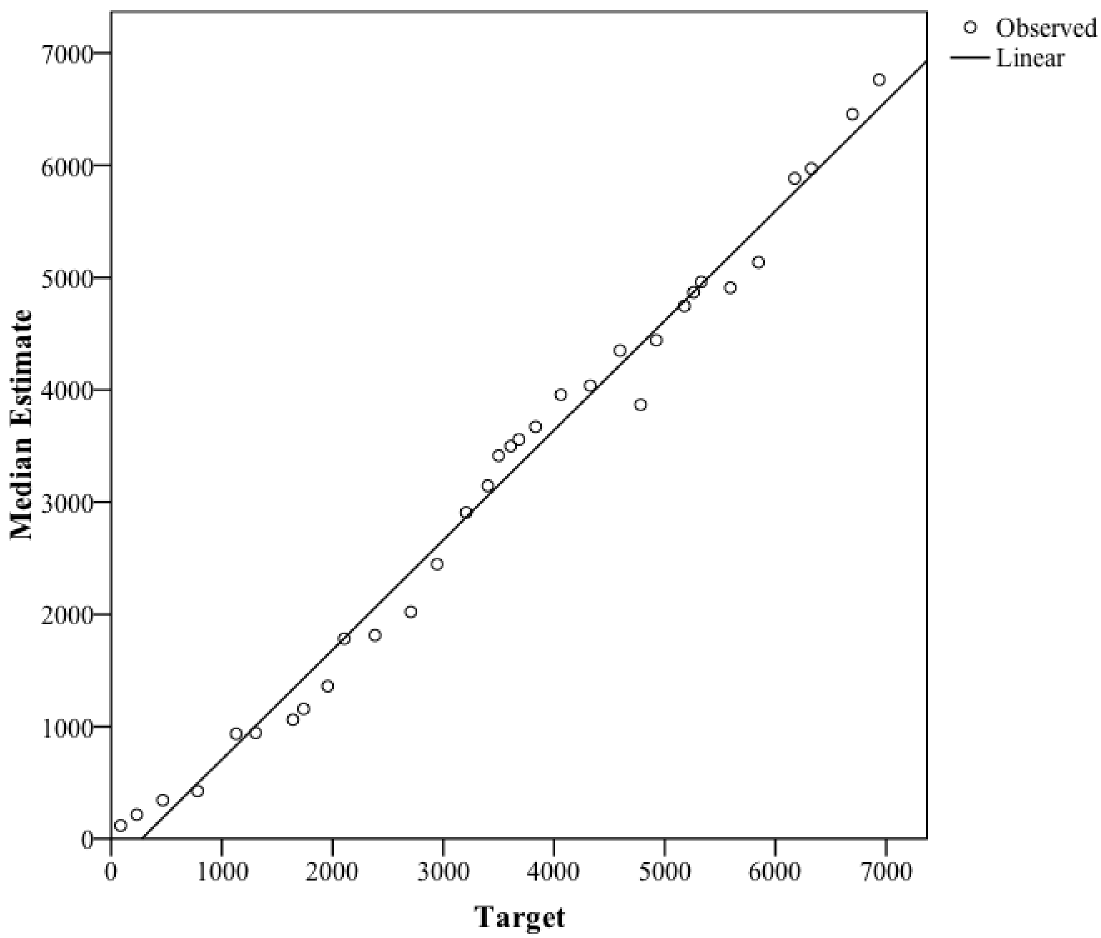

Figure 8. Linear model fit to adult participants' estimates on a 0-7000 number line (R^2=.99).

In the analyses described below, I conducted several specific analyses to test the hypotheses discussed in the Introduction. I used Sullivan et al. (2011) as a guide for some of the analyses, however, I also used findings from Newman et al. (2017) to define questions of interest. As expected, adult participants in the present study demonstrated a high degree of accuracy when placing their estimates. Similar to Newman et al. (2017), the data was well fit by a linear model (98.7%). Despite the high degree of linearity overall, Figure 8 shows that median estimates deviate from a perfect linear fit. Therefore, mean PAE scores were used to further investigate adults' number line estimation performance.

Multi-digit number processing

To test the first two hypotheses, regions of interest 50 pixels wide and 104 pixels tall were defined for each digit in the target number to determine where attention was allocated during multi-digit number processing. All targets were presented as four digits (e.g., 0086). Sullivan et al. (2011) considered fixations and dwell times as "total processing resources". Therefore, both measures were considered in this analysis.

A one-way ANOVA was used to examine mean first fixation frequency to test my first hypothesis that the thousand digit would receive the most first fixations. As shown in Figure 9, there was a significant difference between target digit regions for the first fixation, $F(1.865, 91.40)=41.79$, $p < .001$, $\eta_p^2=.460$. Pairwise comparisons showed that the thousands (M=1.08, SD=1.69) and units (M=4.62, SD=5.84) regions each differed from all other regions. The decades (M=12.72, SD=5.58) and hundreds (M=11.46, SD=7.07) regions did not significantly differ from each other. Contrary to my hypothesis, participants were most likely to look at the decades and hundreds digits first and least likely to look at the thousands digit first. This pattern suggests that, in general, the first fixation on the presented target number was in the centre of the number. It is possible that there was a bias for participants to first fixate on the decades and units digits because they had to saccade from the gaze-contingent box on the right to the target number on the left for each trial (see Figure 24). Although this may have resulted in an increase number of first fixations to target digits on the right side of the target number, the substantial amount of attention allocated to the hundreds and decades digits was unexpected.

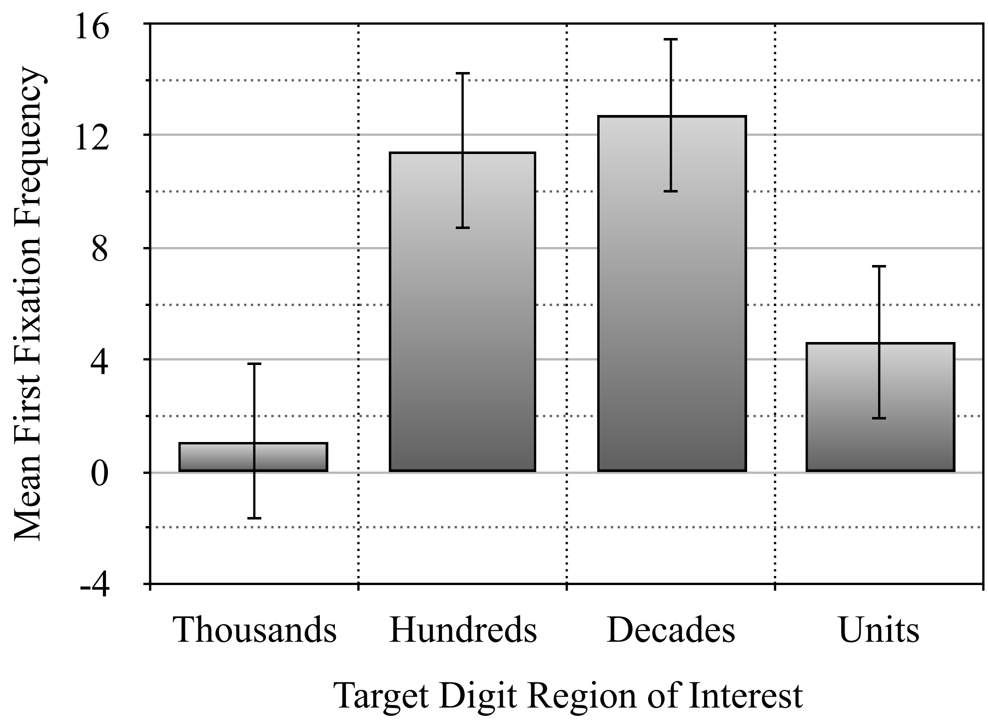

Figure 9. Mean first fixation frequency for each target digit.

Note. Error bars are standard errors of the means.

My second hypothesis stated that the majority of fixations would be in the thousands and hundreds regions. To test this hypothesis, mean fixation counts and dwell times were analyzed in separate one-way ANOVAs to establish if there were differences among target-digit regions. As displayed in Table 2 and Figure 10, there were statistically significant differences in fixations, $F(2.126, 104.154)=103.04$, $p < .001$,

η_p^2=.678, and dwell times, $F(1.853, 90.79)=88.55$, $p < .001$, η_p^2=.644, among target-digit regions.

Table 3

Descriptive statistics of fixations and dwell times for target digit regions of interest

	Fixations		Dwell Time (ms)	
Region	*M*	*SD*	*M*	*SD*
Thousands	.14	.22	38	62
Hundreds	1.08	.49	270	142
Decades	.64	.32	119	83
Units	.16	.18	21	26

N=50.

As shown in Figure 10, pairwise comparisons revealed that the hundreds and decades digit regions significantly differed in fixations and dwell time from the thousands and units regions, as well as from each other. Participants spend the most time looking at the hundreds digit, followed by the decades digit. Since the value of the first digit (thousands) is highly correlated with first fixation accuracy, it is interesting to see significantly more attention paid to other, more central, digits. These results support the "chunking hypothesis" presented by Meyerhoff et al. (2012) which proposes that four-digit numbers are decomposed into chunks which are then processed in parallel. In this case, one chunk would contain the thousands and hundreds digits and the other chunk would contain the decades and units digits.

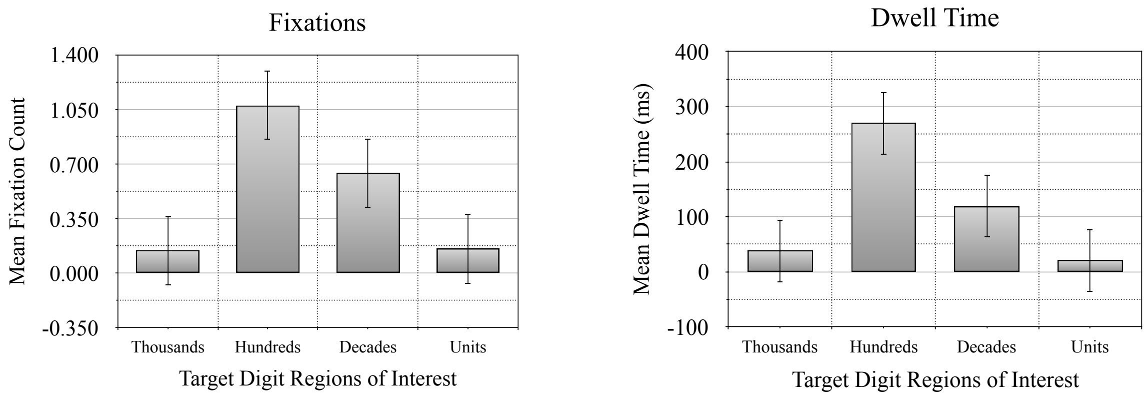

Figure 10. Fixations and dwell times for target-digit regions of interest.

Note. Error bars are standard errors of the means.

I was unable to test whether back-and-forth looking behaviour was related to PAE because participants infrequently looked back at the target number. Specifically, participants looked back at the target number on only 7.7% of trials. The results presented here do not support any of my hypotheses thus far.

Relations of fixations and dwell times to target number

The next analysis was conducted to determine where attentional resources were allocated while completing the number line task. Sullivan and colleagues (2011) found that participants devoted the majority of “total processing resources” (measured by fixation counts and dwell times) to the correct region around a target number. For my third hypothesis, I expected most eye movements to be directed toward the correct region. For hypothesis four, I expected more fixations in the underestimated regions for targets above the midpoint.

Sullivan et al. defined three target regions for each target - underestimated, correct, and overestimated – with each region containing 5% of the number line scale.

For the correct region, the target number was located in the centre of the region of interest with 2.5% of the scale on either side (see Figure 11). In the present research, target regions were similarly defined. Across all participants, 49% of fixations were located in these regions of interest. Mean fixation counts and dwell times were analyzed in separate 2 (location: below the midpoint, above the midpoint) x 3 (region: underestimated, correct, and overestimated) repeated measures ANOVAs using all 33 number line trials.

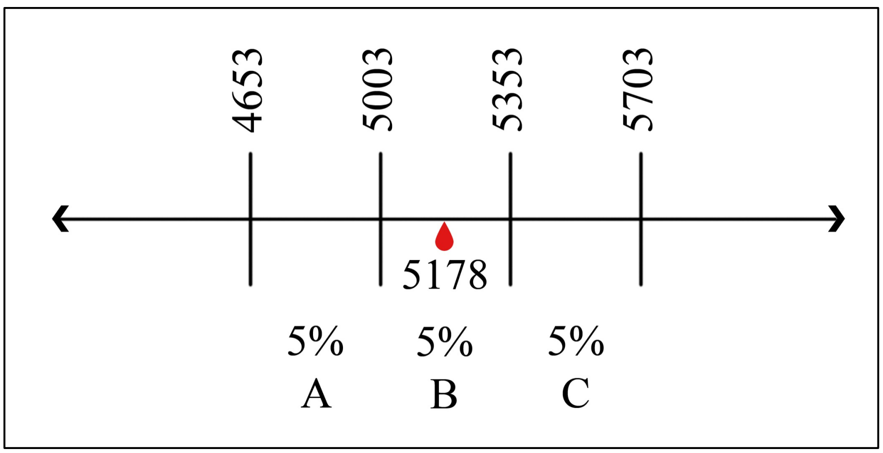

Figure 11. Three target regions of interest: underestimated (A), correct (B), and overestimated (C). For this example, the range of values below the correct region was 4653 – 5003; above the correct region was 5353 – 5703.

Table 4

Descriptive statistics for fixation and dwell data by target region

	Fixation Count (n)		*Dwell times (ms)*	
Region	*M*	*SE*	*M*	*SE*
Underestimated	1.13	.10	630	63
Correct	1.24	.15	619	57
Overestimated	0.64	.10	303	42

There was a significant main effect of region on mean fixation count and dwell time [fixations, $F(2, 98)=18.96$, $p < .001$, $\eta_p^2=.279$; dwell time, $F(1.566, 76.743)=21.25$, $p < .001$, $\eta_p^2=.302$], indicating that the distribution of processing resources differed among regions. Pairwise comparisons showed differences in the allocation of total processing resources between the overestimated region and the underestimated and correct regions (see Table 3 and Figure 12). As shown in Figure 12, participants made more fixations and looked longer at the correct and underestimated target regions compared to the overestimated region. This finding does not support my hypothesis that the majority of resources would be directed towards the correct region since there was no significant difference in processing resources for the underestimated and correct regions. The hypothesis that participants would look more at the underestimated region for targets greater than the midpoint was also unsupported.

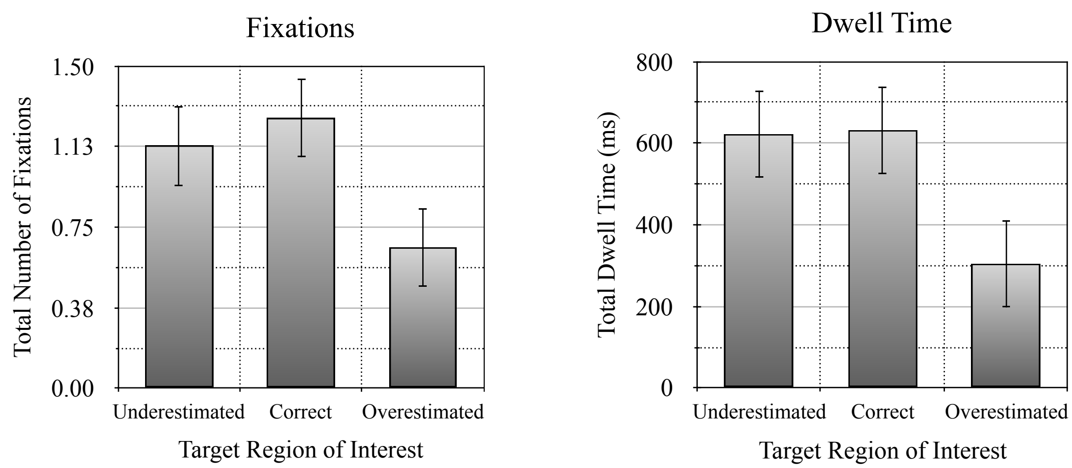

Figure 12. Total fixation counts and dwell times for target regions of interest.

Note. Error bars are standard errors of the means.

Relations between first fixation and estimate accuracy

Sullivan et al. (2011) found that the accuracy of the first fixation on the number line was highly correlated with the accuracy of the estimate. The results of the present study are consistent with this finding. Mean PAE scores were calculated for participants' first fixations and estimates (see Figure 13). First fixation PAE was significantly related to estimate PAE, $r(50) = .75$, 95% BCa CI [.59, .88], $p < .001$. Analyzed separately, first fixation and estimate PAEs were also correlated for targets below and above the midpoint, $r(50)=.72$, 95% BCa CI [.54, .86], $p < .001$, and, $r(50)=.58$, 95% BCa CI [.36, .79], $p < .001$, respectively. These correlations did not significantly differ.

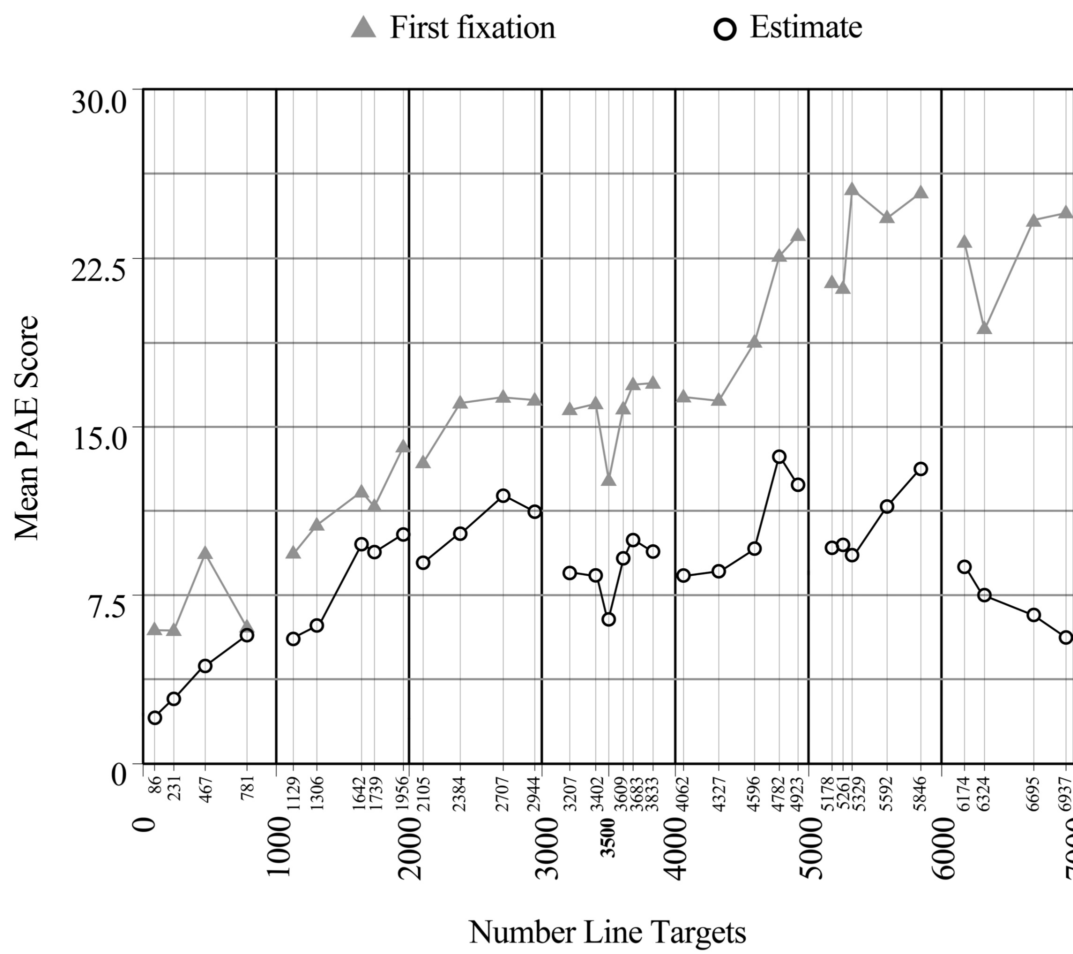

Figure 13. Mean PAE scores for first fixations and estimates for all number line targets.

As shown in Table 4 and Figure 13, there were significant differences between mean PAE scores for both fixation types when the lower and upper halves of the scale were considered separately. A 2 (location: below midpoint, above midpoint) x 2 (fixation type: first or estimate) repeated measures ANOVA was conducted to test my fifth hypothesis that the relationship between fixation types depends on target location, and sixth hypothesis that first fixations have higher PAE scores above the midpoint. As per

Sullivan et al. (2011), targets located within 5% of endpoint and midpoint regions (9) were removed from analysis.

Table 5

Descriptive statistics for fixation type PAE separated by target location size

	Targets below midpoint		Targets above midpoint	
	M	*SD*	*M*	*SD*
First fixation PAE	12.52	4.40	21.43	9.10
Final fixation PAE	8.48	4.47	10.27	6.78

N=50.

Significant main effects were found for fixation type (first, *M*=16.97, *SD*=.86; estimate, *M*=9.38, *SD*=.71), *F*(1, 49)=169.84, $p < .001$, η_p^2=.776, and location (below midpoint, *M*=10.50, *SD*=.59; above midpoint, *M*=15.85, *SD*=1.00), *F*(1, 49)=56.05, $p < .001$, η_p^2=.534, indicating that participants were more accurate with their estimate than their first fixation and that error scores were higher for fixations above the midpoint. Figure 14 displays the significant interaction between fixation type and location, *F*(1, 49)=35.99, $p < .001$, η_p^2=.423, indicating that the relationship between first fixation and estimate accuracy differed by target location. Participants responded less accurately to the locations above the midpoint for both the first fixation and estimate. These results suggest that error increases for both fixation types as the value of the target number increases, resulting in differences in performance for the lower and upper halves of the atypical scale. These results support hypotheses five and six.

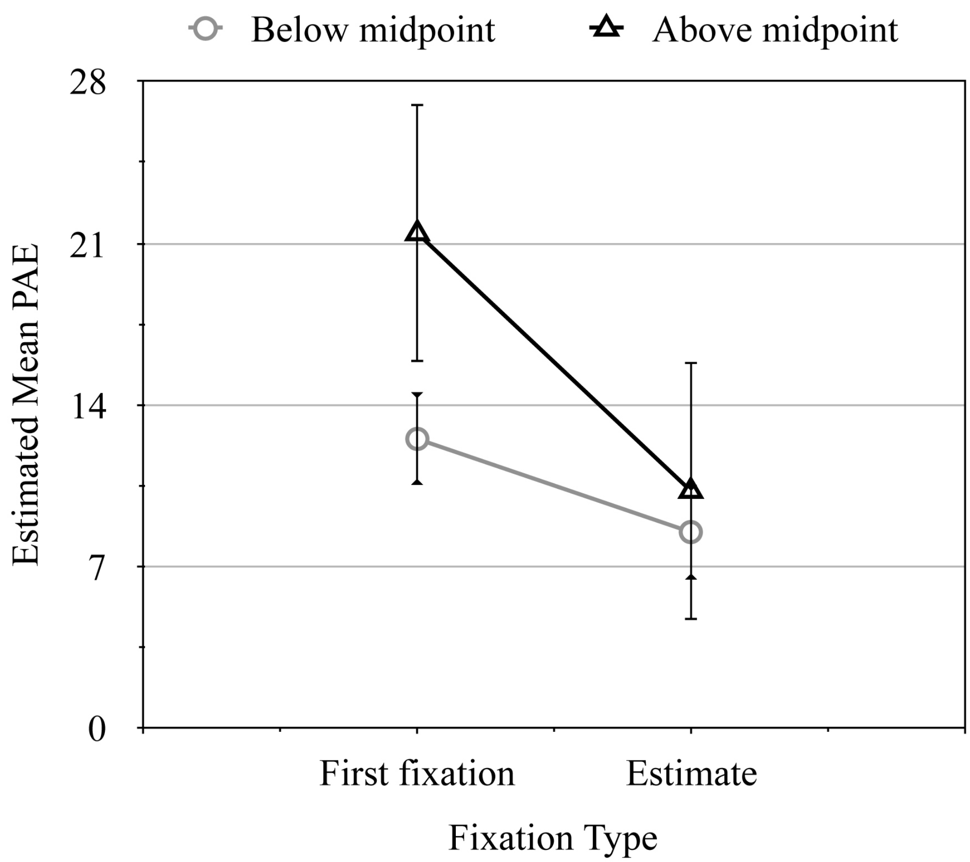

Figure 14. Interaction of target location and fixation type.

Note. Error bars are standard errors of the mean.

Classification of self-reported strategies

Immediately after completing the 33 number line trials, participants were asked to report any strategies used (see Appendix A for post-task questionnaire). As in Newman et al. (2017), strategy groups were classified as either simple or complex based on the number of implicit reference points reported. Participants who reported using a single reference point, or set of spatially equivalent reference points (e.g., midpoint or thousands; see Figure 15), were classified as simple strategy users (*n*=31). Participants who reported using the midpoint in combination with an additional set of spatially equivalent reference points (e.g., the midpoint and thousands; see Figure 15) were classified as complex strategy users (*n*=19). The midpoint and endpoints strategy was classified as a simple strategy because only one implicit reference point was reported. Four possible reference point strategies are shown in Figure 15.

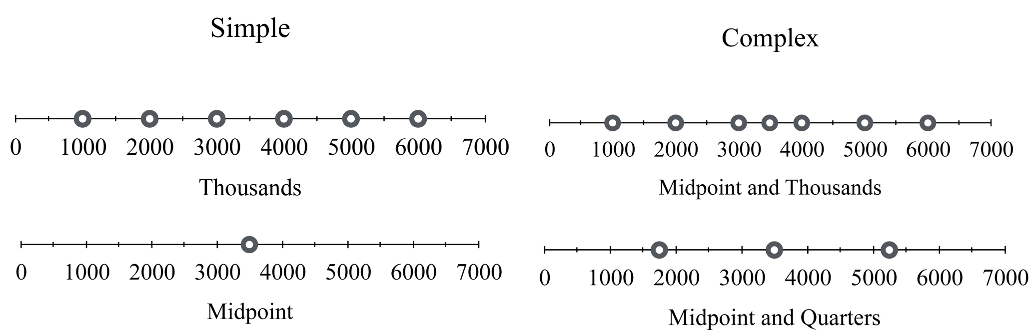

Figure 15. Examples of simple and complex reference point strategies.

As outlined in Table 6, a variety of reference point strategies were reported. Participants belonging to the same strategy group may have reported using different specific strategies. To test hypothesis seven, that specific strategies within a classification group do not significantly differ in error, PAE across strategies were compared in separate one-way ANOVAs. These analyses supported my hypothesis by indicating that overall mean PAE score did not significantly differ among strategies belonging to the simple strategy group, $F(4, 26)=1.55$, $p=.218$, or those belonging to the complex strategy group, $F(4, 14)=.79$, $p=.552$.

Table 6

Frequency of reference point strategies reported by strategy group

Strategy Group	Reference point(s) reported	*n*	Frequency (%)
Simple	Midpoint	14	28
	Midpoint and endpoints	8	16
	Endpoints	5	10
	Thousands	4	8
Complex	Midpoint and quarters	10	20
	Midpoint and thousands	9	18

N=50.

Table 6 shows that complex strategy users had lower mean PAE scores overall and for the lower and upper halves of the number line scale in contrast to simple strategy users.

Table 7

Descriptive statistics for PAE below midpoint and above midpoint separated by group

	Mean PAE < 3500		Mean PAE > 3500		Mean PAE all trials	
	M	*SD*	*M*	*SD*	*M*	*SD*
Simple	8.60	4.45	11.88	7.31	10.25	5.60
Complex	6.14	2.41	6.16	2.78	6.01	2.15

Note. *Simple*, *n*=31; *Complex*, *n*=19.

Figure 16 shows the pattern of accuracy across targets for the two strategy groups. The two groups appear to be equally accurate from targets 86 to 2105; for all other targets, complex strategy users showed less error in their estimates than simple strategy users.

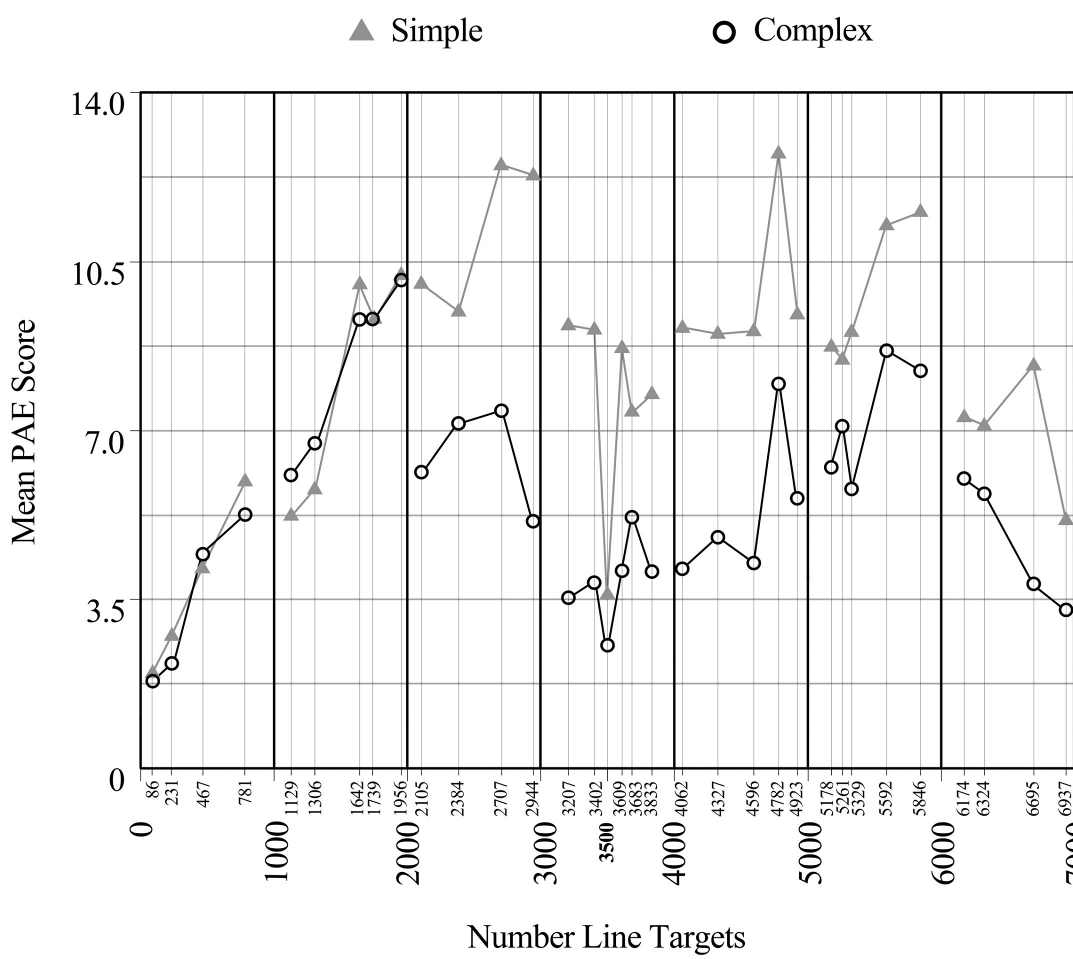

Figure 16. Estimate PAE scores separated by strategy group.

A 2 (location: below midpoint, above midpoint) x 2 (group: simple, complex) mixed ANOVA revealed a significant difference in mean PAE between locations, $F(1, 48)=6.02$, $p=.018$, $\eta_p^2=.112$, and for strategy group, $F(1, 48)=9.82$, $p=.003$, $\eta_p^2=.170$. Participants were more accurate when placing estimates for targets below the midpoint and complex strategy users were more accurate in their estimates than simple strategy users. A significant interaction between location and strategy group was also observed, $F(1, 48)=5.90$, $p=.19$, $\eta_p^2=.109$, indicating that the effect of location on error depends on strategy group. Figure 17 illustrates the difference in mean PAE scores for the two strategy groups and supports hypothesis eight, that strategy group is related to lower error scores for targets above the midpoint. Complex strategy users do not show a difference in errors for locations below and above the midpoint, whereas simple strategy users are less accurate for targets above the midpoint.

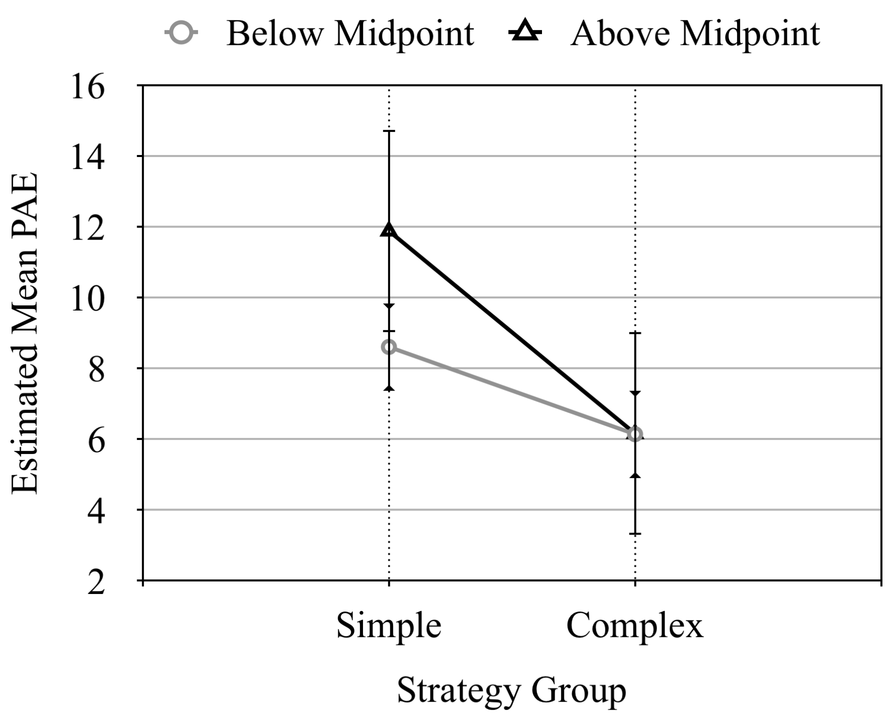

Figure 17. Interaction between strategy group and target location on PAE score.

I also investigated whether mean PAE differed between strategy groups for low-thousand and high-thousand targets to test hypothesis nine. Due to the decrease in mean PAE scores observed for targets located near the endpoints and the midpoint, these 1000-unit intervals were excluded from the analysis. These intervals were also excluded from the analysis conducted in the preliminary study (Newman et al., 2017). Mean PAE scores were calculated for low-thousands targets (e.g., 1000-1500, 2000-2500, 4000-4500, and 5000-5500) and high-thousands targets (e.g., 1500-2000, 2500-3000, 4500-5000, and 5500-6000).

Table 8

Descriptive statistics for low- and high-thousands PAE separated by group

	Low-thousand targets		High-thousand targets	
	M	*SD*	*M*	*SD*
Simple	8.54	4.37	11.09	4.94
Complex	6.09	2.36	7.79	3.13

Note. *Simple*, *n*=31; *Complex*, *n*=19.

A 2 (interval location: low-thousand, high-thousand) x 2 (group: simple, complex) mixed ANOVA revealed a significant difference in mean PAE between interval location, $F(1, 48)=22.26$, $p < .001$, $\eta_p^2=.317$, and strategy group, $F(1, 48)=6.92$, $p=.011$, $\eta_p^2=.126$. Participants had lower error scores for low-thousands targets compared to high-thousands targets and complex strategy users were more accurate than simple strategy users. An interaction between interval location and strategy group was not significant; both strategy groups had more error for high-thousands targets compared to low-thousands targets. The shared pattern of increased error for high-thousands targets, shown in Figure 18, does not support hypothesis nine, that more complex strategy use is related to a lower error score for high-thousands targets compared to simple strategy users.

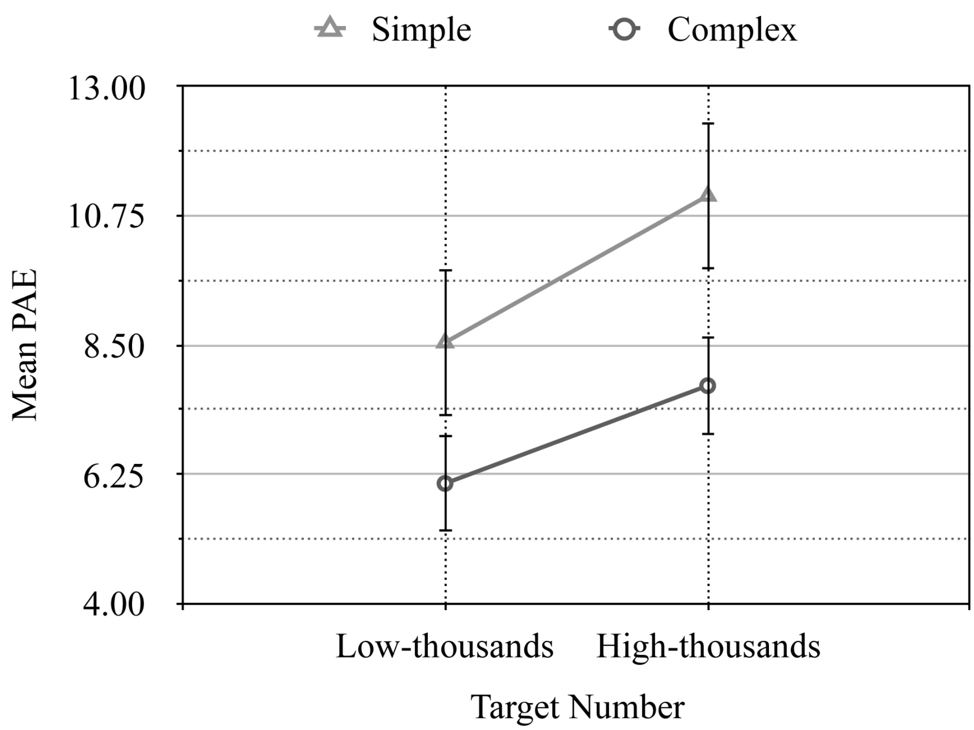

Figure 18. Shared pattern of increased error for high-thousands targets.

Reference point fixations

Previous studies using typical number line scales (e.g., 0-1000) have shown an increase in the number of fixations around bounded endpoints and the midpoint (Schneider et al., 2008; Sullivan et al., 2011). To analyze fixation frequency across the entire scale, 250-unit regions of interest were defined from the 0 endpoint to the 7000 endpoint (28 regions total). As displayed in Figure 19, the W-pattern of fixations observed by Sullivan et al. (2011) was not observed in the current study using an atypical (0-7000) number line scale. These results suggest that the reference point strategies used on an atypical number line differ from those used on a typical number line.

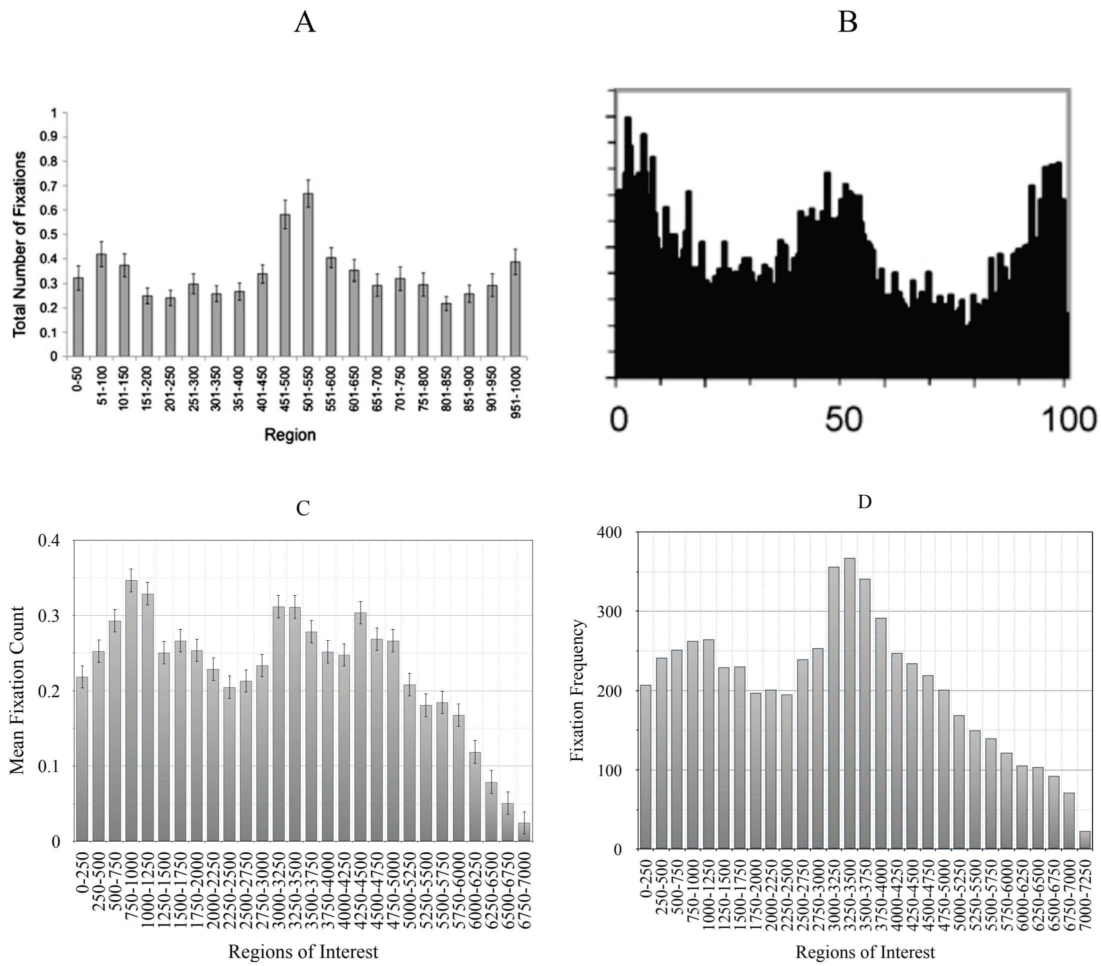

Figure 19. Fixation patterns observed for typical (A & B) and atypical (C & D) number line scales.

Note. Mean fixation count for Sullivan et al.'s 0-1000 scale (2011; A), fixation frequency for Schneider et al.'s 0-100 scale (2008; B), mean fixation count for current 0-7000 scale (C), and total frequency of fixations for current 0-7000 scale (D). Regions of interest in C and D are 250-units. Error bars are standard errors of the means.

As shown in Figure 19D, fixation frequencies decrease as the scalar values increase for values larger than the midpoint. This decline in fixation frequency is especially noticeable when the two halves of the scale are considered separately, as shown in Figure 20.

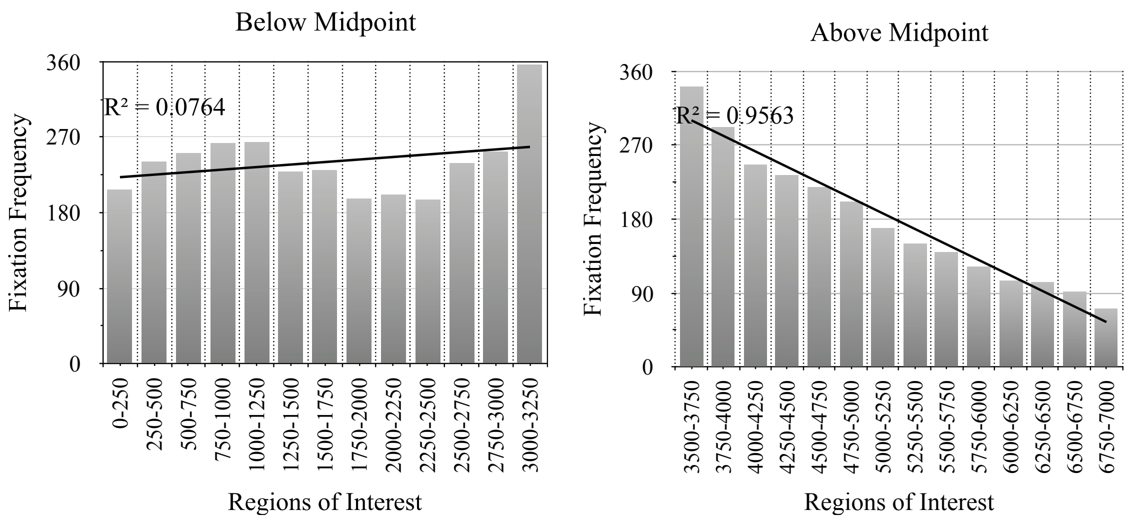

Figure 20. Fixation frequencies for 250-unit regions below and above the midpoint.

Note. *N*=50.

However, graphing all participants together obscures the differences in fixation patterns observed between strategy groups. Figure 21 illustrates the differences in mean fixation count for simple and complex strategy users. Two analyses were conducted to test hypothesis ten, that complex strategy users have more fixations overall and more fixations located in reference point regions compared to simple strategy users. Simple strategy users (*M*=5.35, *SD*=2.03) had fewer overall fixations on the number line compared to complex strategy users (*M*=7.83, *SD*=4.22), *t*(48)=-2.40, *p*=.025, *d*=.75. This is especially noticeable between the 3000 and 6000 locations on the scale.

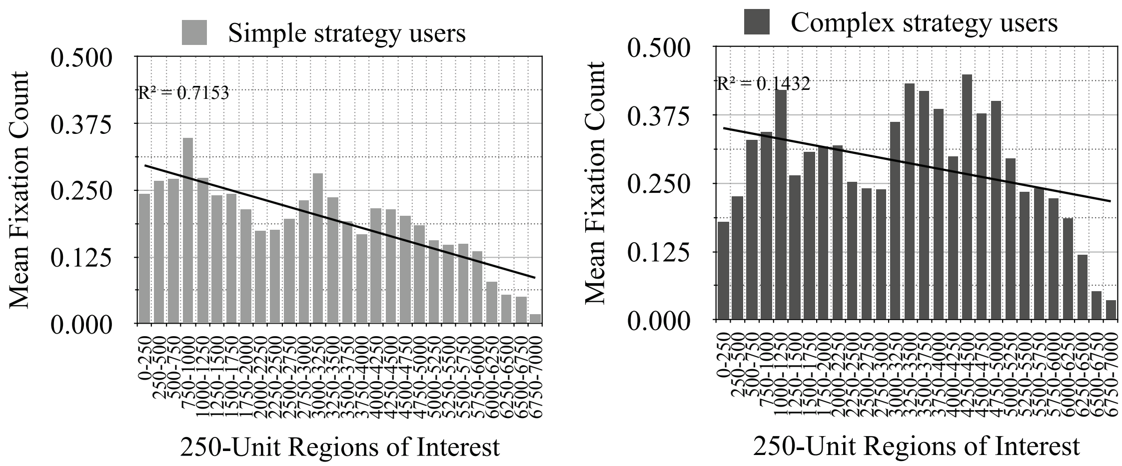

Figure 21. Mean fixation counts for regions of interest by self-reported strategy group.

The contrast between fixation patterns suggests that the strategy groups employ different reference point strategies when locating targets in the lower and upper halves of the scale. To statistically examine this, nine 500-unit reference point regions of interest were defined for the endpoints, midpoint, and thousand values. Reference point values were located in the centre of the region of interest with 250 units (3.6%) of the scale on either side (e.g., 1000: 750-1250; 7000: 6750-7250). Mean fixations were analyzed in a 2 (group: simple, complex) x 9 (reference point: 0, 1000, 2000, 3000, 3500, 4000, 5000, 6000, 7000) mixed ANOVA. The midpoint target was removed as a potential confound leaving 32 targets to be analyzed.

As displayed in Figure 22, there was a significant main effect of region on mean fixation count, $F(5.851, 219.107)=19.80$, $p < .001$, $\eta_p^2=.292$. Pairwise comparisons showed differences in total processing resources between the 7000-region and all other regions with the exception of the 0 endpoint. The 6000 region also differed from all other regions except the 0 endpoint. The 5000 region differed from all regions with the exception of the 0 endpoint and the 2000 region. Differences were also found between the 0 endpoint and the 1000, 3000, 3500, and 4000 regions. Lastly, the 1000 region was significantly different from the 2000 region. As shown in Figure 22, people looked most at the midpoint, 1000, 3000, and 4,000 regions, and less at the endpoints, 2000, 5000, and 6000 regions.

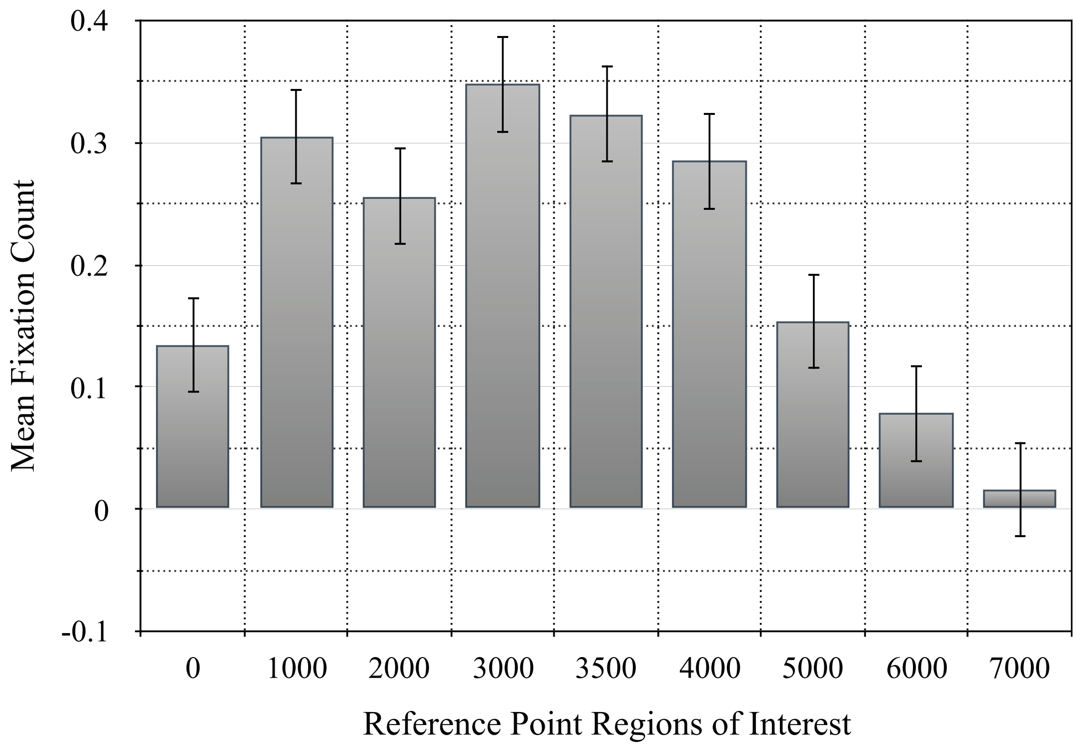

Figure 22. Mean fixation counts for reference point regions of interest.

Note. Error bars are standard errors of the means.

There was also a significant interaction between strategy group and reference point region, $F(4.565, 219.107)=3.5$, $p=.006$, $\eta_p^2=.68$, as shown in Figure 23. Complex strategy users had more mean fixations for reference points from 3500 onward, but the only significant difference between strategy groups was found in the 4000 reference point region, $t(48)=-3.08$, $p=.005$, $d=.97$ (Bonferroni corrected p values used). These findings support hypothesis ten.

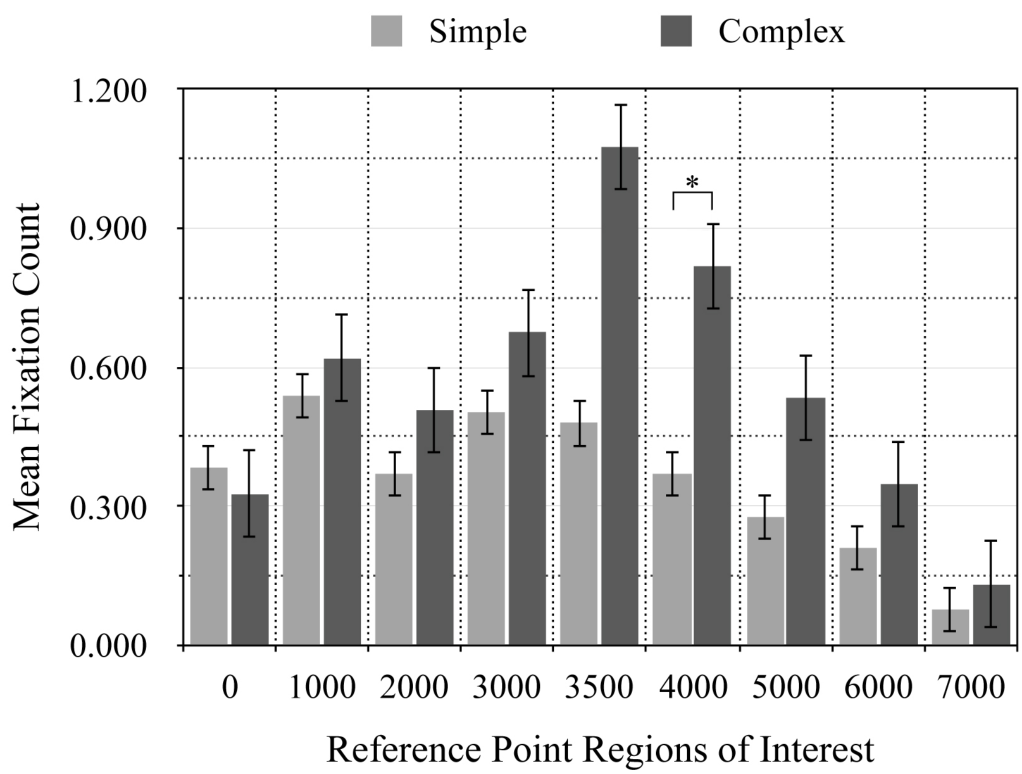

Figure 23. Interaction between strategy group and reference point regions of interest.

Note. * $p=.005$

In summary, simple strategy users looked less at regions greater than the midpoint than did complex strategy users. This pattern may be unique to atypical number line scales. Complex strategy users have the largest number of fixations in the midpoint region, and visited the 4000 reference point region more than simple strategy users.

Figure 24 illustrates a common pattern of eye movement behaviour for a target located below the midpoint. In this example, a simple strategy user looked at target 2384, overestimated the target's location with their first fixation to the number line, and then used the 2000 and 3000 markers to adjust their estimate. A different pattern was observed for targets located above the midpoint, as shown in Figure 25. For target 5846, a complex strategy user first looked at the midpoint, then adjusted upward, fixating near the 4000 and 5000 markers, until placing their estimate. The fixation data examples shown in Figures 24 and 25 support the view that participants used thousands reference points for targets below the midpoint, but anchored and adjusted from the midpoint for targets located at the upper end of the scale. This general approach is consistent with the overall fixation patterns which show (a) decreased attention to the upper half of the scale for simple strategy users, with very little attention to the upper endpoint by either strategy group, (b) relatively more attention to locations near the left endpoint (i.e., 1000 location), and (c) more attention to locations around and just beyond the midpoint (i.e., 3000, 3500, and 4000 locations) for complex strategy users. The fixation patterns observed suggest that participants appear to have anchored and adjusted from left-to-right from the low endpoint for targets below the midpoint. The eleventh hypothesis, that participants would be more likely to use thousands reference points for an atypical scale was partially confirmed; only complex strategy users appear to make use of the thousands reference points for targets above the midpoint.

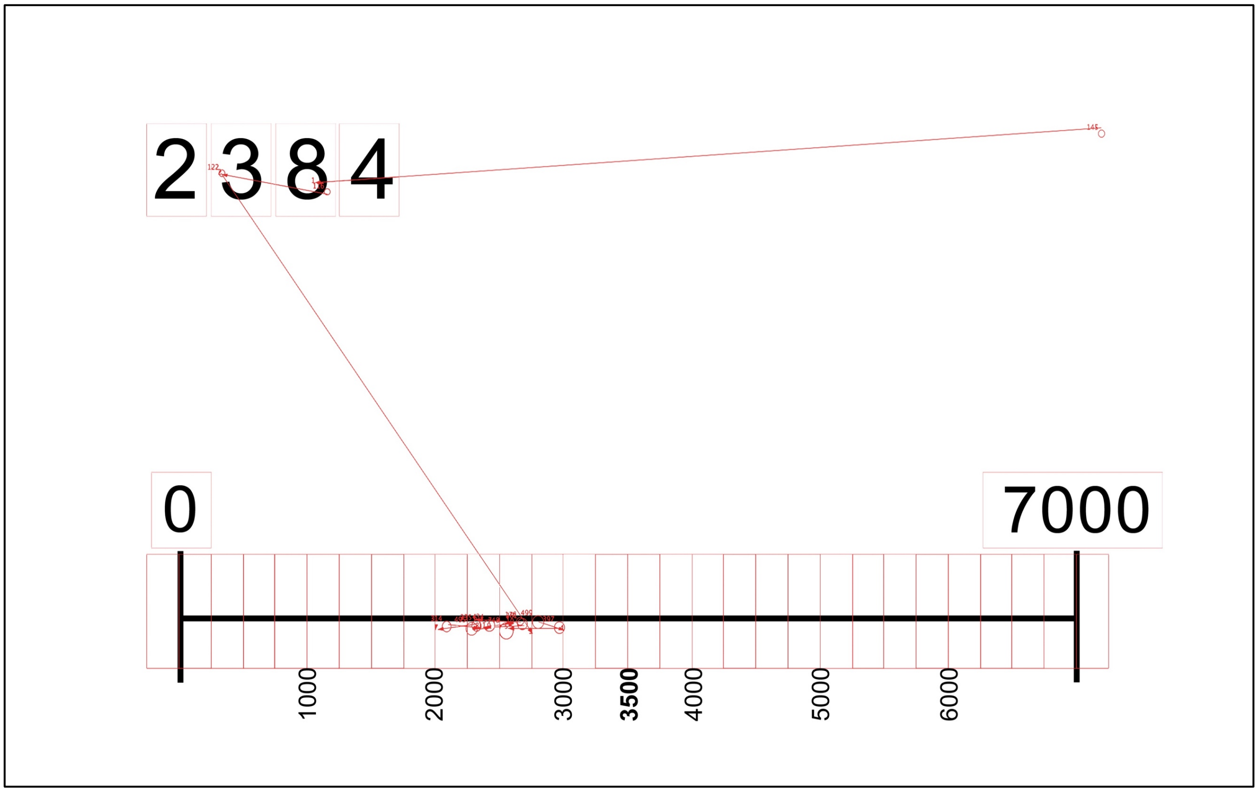

Figure 24. Example of reference point fixations for a target below the midpoint by a simple strategy user.

Note. Thousands and midpoint shown for illustration purposes only.

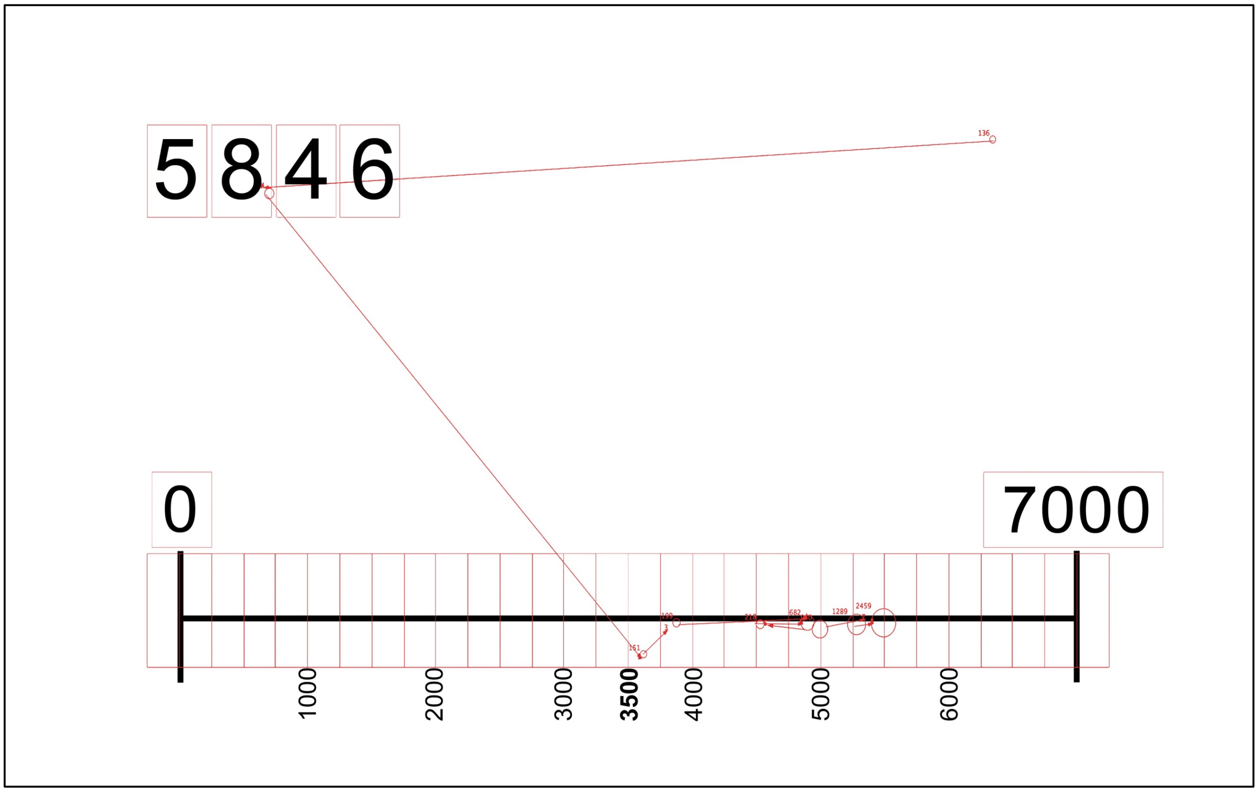

Figure 25. Example of reference point fixations for a target above the midpoint by a complex strategy user.

Note. Thousands and midpoint shown for illustration purposes only.

Predictors of mean PAE for lower and upper regions of the number line scale

The previous analysis revealed distinct fixation patterns for simple and complex strategy users. The following analyses will investigate factors that account for individual differences in performance for the lower and upper halves of the atypical number line scale. Considering the earlier finding that target location had a significant main effect on mean PAE, the following analysis was done to identify predictors of error for the lower and upper regions of the scale.

Table 9

Descriptive statistics for reference point fixations below and above the midpoint by strategy group

	Simple		Complex	
	M	*SD*	*M*	*SD*
Reference point fixations below midpoint	.45	.22	.53	.29
Reference point fixations above midpoint	.23	.12	.46	.31

N=50.

The question of interest was whether participants' accuracy of locating targets below the midpoint (i.e., their PAE for targets below 3500) and above the midpoint (i.e., their mean PAE for targets above 3500) was related to (a) their fraction skill, as indexed by their performance on the six fraction questions on the BMA-3R, (b) their self-reported strategy, and (c) the frequency with which they used implicit reference points. As shown in Table 9, mean fixation counts were calculated for reference point targets below the midpoint (0, 1000, 2000, 3000) and above the midpoint (4000, 5000, 6000, 7000) for each participant and considered final predictors in the regression analysis. Correlations among these measures are shown in Table 10 for targets below the midpoint and Table 11 for targets above the midpoint.

Table 10

Correlations for estimate mean PAE for targets below the midpoint

Variables	1	2	3
1. Mean final PAE for targets < 3500	-		
2. Fraction skill	-.526**	-	
3. Strategy group[a]	-.305*	.187	-
4. Mean Reference points used < 3500	.273	-.198	.164

Note. $^{*}p < .05$, $^{**}p < .001$.

[a]Strategy: *1=simple, 2=complex.*

As shown in Table 10, participants' accuracy on targets below the midpoint was related to higher fraction skill. Accuracy was also related to strategy group with complex strategy users demonstrating less error than simple strategy users. The number of reference point fixations visited was not significantly correlated with mean PAE for targets below the midpoint. As shown in Table 11, participants' accuracy above the midpoint was also related to higher fraction skill and complex strategy use. In addition, the number of reference point fixations above the midpoint was correlated with mean PAE. Thus, performance on targets above the midpoint was linked to participants' use of reference points in that range.

Table 11

Correlations for estimate mean PAE for targets above the midpoint

Variables	1	2	3
1. Mean final PAE for targets > 3500	-		
2. Fraction skill	-.531*	-	
3. Strategy group[a]	-.426*	.187	-
4. Mean Reference points used > 3500	-.493*	.205	.464*

Note. $*p \leq .001$.

[a]Strategy: *1=simple, 2=complex.*

I conducted two hierarchical linear regressions for hypotheses twelve and thirteen to test whether the relations between accuracy and the predictors reflected shared or unique variance.

The regression analysis for targets below the midpoint supported hypothesis twelve, that mean PAE score was predicted by fraction skill (Model 1) and self-reported strategy (Model 2). The final model including fraction skill, strategy group, and reference point fixation data was significant, $F(3, 46)=9.01$, $p < .001$. In this model, fraction skill and self-reported strategy were significant predictors (see Table 11).

Table 12

Hierarchical multiple regression predicting mean estimate PAE for targets smaller than the midpoint from fraction skill, strategy group, and reference point fixation data

	Mean estimate PAE for targets smaller than 3500					
	Model 1		Model 2		Model 3	
Variable	B	β	B	β	B	β
Constant	11.69***		13.78***		12.31***	
Fraction skill	-1.41***	-.53***	-1.30***	-.49***	-1.15***	-.43***
Strategy group[a]			-1.73	-.21	-2.12*	-.26*
Ref pts used < 3500					3.68	.23
R^2	.277***		.321***		.329***	
F change in R^2	18.39***		3.06		3.57	

Note. *$p < .05$, ***$p \leq .001$.

[a]Strategy: *1=simple, 2=complex.*

As shown in Table 13, the frequency of reference point fixations above the midpoint predicted accuracy for targets in this region independently of the other predictors, $F(1, 46)=6.45$, $p=.015$, supporting hypothesis thirteen. The final model including fraction skill, self-reported strategy, and reference point fixations was significant. However, self-reported strategy was not a significant unique predictor in the final regression model. This is likely due to the shared variance with the reference point fixation data which accounted for 7.5% of unique variance after controlling for fraction skill. Participants who were more accurate above the midpoint had higher fraction skill, were more likely to be complex strategy users, and used more reference points above the midpoint. All three of these predictors showed similar correlations with PAE above the midpoint.

Table 13

Hierarchical multiple regression predicting mean estimate PAE for targets greater than the midpoint from fraction skill, strategy group, and reference point fixation data

	Mean estimate PAE for targets larger than 3500					
	Model 1		Model 2		Model 3	
Variable	B	β	B	β	B	β
Constant	16.45***		21.93***		21.66***	
Fraction skill	-2.36***	-.53***	-2.08***	-.47***	-1.91***	-.43***
Strategy group[a]			-4.55**	-.34**	-2.71	-.20
Ref pts used > 3500					-8.70*	-.31*
R^2	.282***		.393***		.468***	
F change in R^2	18.90***		8.57**		6.45*	

Note. *p < .05, **p=.005, ***$p \le .001$.

[a]Strategy: *1=simple, 2=complex.*

In summary, the number of reference points visited below the midpoint did not predict mean error score for targets below the midpoint. However, the number of reference points visited above the midpoint was related to mean error score for targets above the midpoint; people who fixated on more reference points above the midpoint had lower error scores than people who fixated on fewer reference points above the midpoint. Fraction skill was related to performance on both halves of the scale, however, and strategy group predicted performance on both halves of the scale, although not uniquely above the midpoint when the frequency of reference points used was included in the regression; the reference point fixation data shares variance with strategy group. These results partially support the hypothesis that strategy group and reference point fixations would predict mean PAE for targets above the midpoint. These results also suggest that performance on an atypical version of this estimation task reflects an understanding of proportional reasoning, both in the general sense that fraction knowledge aids in the understanding of the relationship between the target number and the high endpoint, and more specifically in the sense that the use of reference points is related to a decrease in error for the more challenging values located in the upper half of the scale.

Discussion

The present study focused on investigating the role of reference point use in relation to performance on an atypical number line task with adult participants. Two key research questions were of interest: i) do adults adjust their strategy in response to an atypical number line scale, and ii) is the use of implicit reference points related to performance? Previous studies using typical number line scales have shown consistent patterns of reduced error and increased fixations around the midpoint and endpoints (Ashcraft & Moore, 2012; Schneider et al., 2008; Sullivan et al., 2011). These findings suggest that adults and older children use the explicitly marked endpoints and an implicit midpoint to estimate target locations. Contrary to these findings, the patterns of error and reference point fixations presented for the 0-7000 number line suggest that adults vary their reference point strategy in response to the scale used (typical vs. atypical). The presented results show that participants who used fewer reference points had higher error scores compared to participants who used more reference points. This finding was especially evident for the more challenging targets in the upper half of the atypical number line scale. The variation demonstrated within an adult population suggests that adults are not accessing a mental number line representation, but are using different proportional reasoning strategies to translate a numerical value to a physical location.

Much of the research related to reference point use focuses on the use of midpoint and endpoint reference points on typical, base-10 scales. It is likely that these scales are not challenging enough for skilled, adult participants. One reason for this opinion is that often only a single implicit reference point – the midpoint – is required. Typical scales do not allow for the examination of different strategies used by adults. Atypical number line scales not only present adult participants with a more challenging task, but allow for the use of strategies that incorporate different implicit reference points.

In the present study, error was lower for targets located closer to 7000 despite there being few fixations in the 7000 reference point region. Interestingly, many of the trial-level self-reports contradicted this finding. When asked how they located target 6695, 82% (*n*=41) of people reported looking at the 7000 endpoint and working their way backwards down the number line to place their estimate. However, nearly a quarter of people who reported using this strategy (22%; *n*=9) did not have a single fixation in the 6750-7000 region for target 6695 or for any of the other 32 trials. This suggests that participants can utilize the explicitly marked bounded endpoints without having to directly attend to them. Therefore, scales that require implicit reference point use have the potential to be more informative in regard to strategy use. The incongruence between trial-level reports and fixation data makes eye tracking methodology especially relevant for testing adult participants on estimation tasks.

The eye movement data collected here confirmed many of the hypotheses that were based on behavioural data collected by Newman et al. (2017). Reference point fixation data supported the participants' self-reported strategies that complex strategy users visit more implicit reference point regions than simple strategy users when placing their estimates. Consistent with the findings from Newman et al. (2017), strategy group predicted individual differences in performance. Although strategy classification is related to performance, further study is needed to fully understand what simple and complex strategies represent. I suggest that self-reported strategy indicates the participants' approach to the problem and how careful they are when placing their estimates. For example, simple strategy users may approach the task using an heuristic strategy that is quick and consistent, and complex strategy users may adapt their reference point strategy in response to specific targets.

The eye movement data analyzed in the present study supported many of the hypotheses based on the work of Sullivan and colleagues (2011). Namely, fixations to the target are highly accurate and the first fixation on the number line is related to the accuracy of the estimate. The finding that adults allocate the same percentage of processing resources to the target number regions on an atypical scale as they do on Sullivan et al.'s (2011) typical scale suggests that adults are able to adapt to different number line scales.

In an effort to extend Sullivan et al.'s (2011) findings, I found evidence to support my hypothesis that the relationship between first fixation and estimate would depend on whether the target was located below or above the midpoint value of the scale. The interaction between fixation type and location supports my position that targets larger than the midpoint on an atypical number line scale present a greater challenge for participants in comparison to targets smaller than the midpoint. The increase in error for targets located between 4000-6000 suggests that targets in this area are more challenging for both strategy groups. This difficulty is likely related to participants' tendency to anchor their first fixation to the left of the target and adjust upwards toward the high endpoint. The challenge of accurately placing an anchor from which to adjust then increases in proportion to the target number's value.

This pattern of anchoring and adjusting from left-to-right was evident when comparing mean error scores for targets below and above the midpoint, but also when comparing the error of low-thousands targets to high-thousands targets. It appears that both strategy groups display a left-to-right estimation pattern, but that the use of additional reference points by the complex strategy group increases the accuracy of their estimates across the entire scale. Instead of error continuing to increase in response to targets above the midpoint, complex strategy users continue to anchor at additional reference points in the upper half of the scale, thereby reducing the amount of error in their estimates.

The absence of an interaction between interval location (low- and high-thousands targets) and strategy group can also be interpreted as support for the link between implicit reference point use and decreased error. It is unlikely that participants use implicit reference points within an interval of 1000 units. Therefore, if participants anchor at a thousand marker (e.g., 4000) and adjust upward in relation to the size of the target within that interval (e.g., 4782), then the increase in error would not significantly differ between groups because complex strategy users would be unable to fixate on additional implicit reference points to help refine their estimate. This finding suggests that the differences in mean error scores observed between strategy groups is related to the use of implicit reference points and not strictly to the size of the target in relation to its scale (e.g., entire 0-7000 range, or 1000-2000 interval). As well, the more challenging nature of the atypical number line scale makes it an appropriate tool to investigate estimation strategies used by adult participants.

Mean error scores and fixation data show that people use the thousands as reference points on an atypical scale. The relatively low mean error scores and high fixation count around the 1000 region suggests that both strategy groups use the 1000 marker to calibrate the size of a 1000-unit interval and estimate upwards from this location. This finding, along with the pattern of estimating from left-to-right, is similar to the patterns observed on the unbounded number line task introduced by Cohen and Blanc-Goldhammer (2011). Figure 26 illustrates the difference between the bounded and unbounded versions of the task. In the unbounded version, participants are presented with a horizontal line showing the size of one unit. Participants are required to count upwards from the unit displayed to locate their estimate. Cohen and Blanc-Goldhammer (2011) developed this version in response to the suggestion that the bounded number lines measure proportional judgment ability and not the mental representation of quantity.

It is thought that the unbounded version of the task is a purer measure of the mental number line.

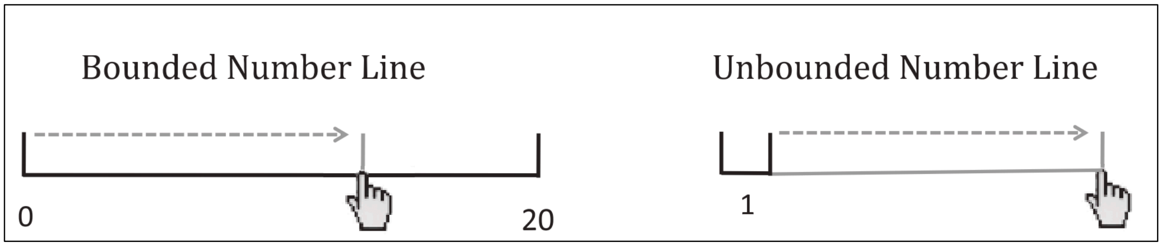

Figure 26. Examples of bounded and unbounded number line tasks (Cohen & Sarnecka, 2014).

If the unbounded version is a more accurate measure of mental representation, then it would follow that performance on the unbounded number line would be related to math ability. However, Link et al. (2014) tested children in grade four on basic numeric and arithmetic measures (i.e., addition, subtraction, multiplication, number completion, number comparison, bounded number line, and unbounded number line) and found that children's unbounded number line performance was not related to any of the math measures tested. As previously demonstrated, the bounded task was related to math ability (i.e., addition, subtraction, and number comparison; Booth & Siegler, 2008; Friso-van den Bos et al., 2015; Link et al., 2014; Siegler & Opfer, 2004). If the unbounded number line task is intended to be a more direct measure of the mental number line, then why is performance not related to math ability? The similar pattern observed for targets below the midpoint on the atypical scale and in the unbounded version of the task requires further study. It is possible that counting upwards from left-to-right by thousands is an effective strategy for smaller targets.

Sullivan et al. (2011) proposed that "preprocessing measures" accounted for the accuracy of the first fixation on the number line, but did not elaborate as to what was meant by this term. First fixations for targets above the midpoint had a higher degree of error in comparison to first fixations below the midpoint. However, the error in participants' estimates did not correspond with the increase in error of their first fixation for targets above the midpoint. Participants were still able to place their estimates in a linear fashion despite the inaccuracy of their first fixation to the number line for high targets. Bearing in mind that fraction knowledge was a significant predictor of error for targets in both the lower and upper halves of the scale, this measure may account for some of the preprocessing referenced by Sullivan and colleagues (2011). Further analyses are needed to evaluate the relationship between fraction knowledge and first fixation accuracy, but it is possible that fraction knowledge is related to the accuracy of the first fixation and that implicit reference point use is related to the degree of error for estimates, especially for targets located above the midpoint.

Limitations

A key limitation of the present study concerns the presentation of the number line trials. The location of the gaze-contingent square in the top right corner and the target number in the top left corner (see Figure 24) may have influenced target digit fixations. As well, due to the location of the target number, participants may have been biased towards underestimation when looking from the target number to the number line. Previous studies, primarily with children, have presented the target number above the centre of the number line (Booth & Siegler, 2006; Heine et al., 2010; Opfer & Siegler, 2007; Siegler & Opfer, 2003; Slusser et al., 2013; Stapel et al., 2015; Thompson & Opfer, 2010). However, the research concerning error and fixation data has demonstrated that older children and adults make use of the midpoint to anchor and adjust their estimates. For this reason, the top left location was chosen to display the target number in an effort to avoid cueing adult participants to the midpoint location. An alternative procedure, used by Sullivan and colleagues (2011), was to present target numbers aurally. Aural presentation would reduce bias in looking direction, but would simultaneously increase demands on working memory and prohibit the study of fixation and dwell data for individual target digits. Although the top left display of target number may have influenced looking behaviour, I felt that the consistency of a fixed location that did not cue participants to the midpoint was the best option (Peeters et al., 2016).

A further limitation involves the use of an estimation task to investigate multi-digit number processing. The number line task is not traditionally used to assess multi-digit number processing and it is possible that the task-specific nature of this estimation task warrants it an inappropriate measure for this kind of investigation. The finding that participants direct their attention towards the centre region of a four-digit number was surprising, but this finding must be interpreted with caution. Additional research into multi-digit number processing beyond two digits is necessary before a theoretical position can be asserted for large numbers. It is likely that numbers of varying lengths are processed differently and further empirical study is needed for numbers of specific lengths. With that in mind, the chunking hypothesis set forth by Meyerhoff et al. (2012) appears to be a good starting position from which to investigate the processing of four-digit numbers.

Implications

The flexibility with which adult participants adjust their strategy in response to the scale used, indicates that further study is needed to investigate factors that account for individual variation. As well, researchers should be cautious about overstating claims regarding what estimation tasks measure and generalizing results between children and adult samples. It is mistaken to assume that all adult participants apply the same estimation strategies to number line tasks simply based on linear performance and low error scores near reference points (most of which are often explicitly marked).

The results presented here strongly support the use of eye tracking methods for studying cognitive processes, such as number line estimation. The high degree of accuracy demonstrated by adults can obscure individual differences in performance when only mean patterns of error are studied. Eye tracking data can be analyzed to study differences in performance in a more nuanced way. For example, the fixation data presented here revealed two key findings: i) participants adjust their looking behaviour in response to the type of scale used (e.g., fewer fixations in the region above the midpoint for an atypical scale), and ii) individuals invoke different reference point strategies when locating their estimates (e.g., complex strategy users have more reference point fixations). The study of eye movement data reduces the need to make inferences about cognitive processes based on behavioural data (i.e., patterns of error) and self-reports. Most importantly, the reference point fixation data presented here demonstrates that reference point use is related to accurate performance on the number line task. Eye movement data can be used to evaluate competing theoretical models regarding the underlying abilities measured. The use of reference points supports the position that the number line estimation task requires number knowledge and proportional reasoning abilities. The variations in performance observed on an atypical version of the task do not support a unified model of mental representation of quantity.

Future Research

Number line research involving adult participants should aim to identify additional predictors of performance since a great deal of individual variability remains unaccounted for. Other tasks that measure landmark or reference point use, such as navigation ability, could be assessed to see if this kind of spatial measure accounts for any shared or unique variance. Gross motor skills could also be assessed to see if there is a relationship between interacting with one's environment, in an embodied sense, and visual estimation performance.

More work is also needed using scales of different magnitudes and layouts. For example, performance could be assessed using scales with highly irregular number ranges (e.g., 5299-7912) or atypical scales that do not have vertical hatch marks explicitly displaying the bounded endpoints. It would be interesting to investigate performance and looking behaviour in response to highly irregular and/or unfamiliar scales.

Although adult participants are very skilled at this task, it is possible that differences in estimation strategies and ability may influence decision making and have an effect on general well-being. The understanding of large magnitudes and the ability to appreciate the relationship between these values is important for making informed decisions related to public policy. For instance, Landy et al. (2013) found that estimators who are less linear in performance are more likely to positively evaluate ineffective political strategies in relation to more accurate, linear estimators. There is a growing body of research identifying a strong association between political affiliation and bias concerning the interpretation of numerical quantity (Landy et al., 2013; Landy et al., 2014; Landy et al., 2017; Olsen, 2013). Estimation biases could affect how individuals make decisions based on numerical information in other areas such as health care and the environment. It is common for medical and environmental data to be presented in large or unfamiliar units of measurement across a variety of different scales. For example, an environmental contaminant may be reported in parts per million for an area measured in square kilometers with a cleanup cost presented in dollars as a percentage of a budget. Processing numerical information across a variety of scales is especially difficult when a reference is not provided. It is worth exploring whether relationships exist between estimation performance, in general, and the ability to make informed economic, health, and environmental decisions based on large or unfamiliar magnitudes.

Conclusion

The purpose of the present study was to use eye tracking data to investigate the relationship between reference point use and accuracy on an atypical number line task. University-aged participants completed 33 trials using a 0-7000 number line task while fixation and duration eye movement data was collected. The results show that self-reported strategy is related to overall error and that reference point use benefits performance on this task for numbers larger than the midpoint value. These results indicate that variations in the highly linear performance demonstrated by adult participants can be manipulated with the use of an atypical number line scale. In conclusion, the relationship between reference point use and accuracy on an atypical number line task supports the notion that adult participants use proportional reasoning skills to locate their estimates. The findings discussed here do not support the use of a mental number line representation to complete the task.

Appendix A

Post-task questionnaire for 0-7000 number line task

EXPERIMENTER: We are very interested in what strategies adults might use to complete a number line task. I am going to ask you some questions about the task and record your answers. There are a number of different ways to do it and there are no right or wrong answers to my questions. Please think about how you completed the task when answering the following questions.

1. How did you decide where a number was located?
2. Did you use any other reference points, either marked (0, 7000) or unmarked, when you were deciding where to place the numbers? If so, what reference points did you use?
3. What was the value of the midpoint of the line?[1]

Let's complete a few trials together. I'd like for you to tell me what you are thinking as you find the location of a number on the line.

4. Can you tell me how you would locate 6695?
    1. Did you use any reference points? If so, which ones?
    2. Did you round the number up or down? If yes, what number did you round to?
5. Can you tell me how you would locate 5178?
    1. Did you use any reference points? If so, which ones?

[1] If the participant did not report the midpoint as part of their reference point strategy, then they were asked to report the midpoint value at the end of the interview.

    2. Did you round the number up or down? If yes, what number did you round to?

6. Can you tell me how you would locate 2707?
    1. Did you use any reference points? If so, which ones?
    2. Did you round the number up or down? If yes, what number did you round to?

Appendix B

Brief Mathematics Assessment - Revised   Participant # ________________

Instructions: Below are some math problems. Pleas complete as many of these problems as you can, answering in the spaces provided. You are allowed to use paper and a pen/pencil to do rough work but please do not use any aids such as calculators. Note that the problems become more difficult as you go along. It is okay to skip questions and leave questions blank. Many people are not able to do all of the problems so please just do your best. You have 10 minutes to complete this task.

1.	2.	3.	4.
$\begin{array}{r} 42 \\ \underline{-21} \end{array}$	$\begin{array}{r} 56 \\ \underline{+17} \end{array}$	$\begin{array}{r} 8 \\ \underline{\times 5} \end{array}$	$\frac{9}{3} =$

5.	6.	7.
$3\frac{1}{2} + 2\frac{1}{2} =$ _____	$\begin{array}{r} 4\frac{1}{4} \\ 3\frac{1}{8} \\ \underline{+\ 2\frac{1}{2}} \end{array}$	$\begin{array}{r} 8\frac{1}{4} \\ \underline{-\ 5\frac{2}{3}} \end{array}$

8.	9.	10.
Write as a percentage. $\frac{1}{8} =$ ______%	Write as a common fraction in lowest terms: $.025 =$ _______	$2^4 =$ ________

11.	12.	13.
$5j - w = 18$ $4\,j - w = 24$ $j =$ ____ $w =$____	$\sqrt{4ab} = 8$ $a =$ __________	Reduce: $\frac{p^2+p}{p^2} * \frac{2p-2}{p^2-1}$ Answer: ____________

[2]

---

[2] The second equation in item 11 was 4j-w=14 in the original BMA-3.

References

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology, 111*, 246-267.

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

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology, 46*(2), 545-551.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*(1), 189-201.

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

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, 13, 83-91.

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

Cohen, D.J., & Sarnecka, B. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology, 50*(6), 1640-1652.

Dehaene, S. (1989). The psychophysics of numerical comparison: a reexamination of apparently incompatible data. *Perception & Psychophysics, 45*(6), 557-566.

Dehaene, S. (1997). *The Number Sense.* Toronto, Ontario, Canada: Oxford University Press.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences, 7*(4), 145-147.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log of linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science, 320*(5880), 1217-1220.

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

Friso- van den Bos, I., Kroesbergen, E., Van Luit, J., Xenidou-Dervou, I., Jonkman, L, Van der Schoot, M., et al. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology, 134*, 12-29.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*(1), 4-15.

Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343-1359.

Geary, D.C., Hoard, M.K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33*, 277-299.

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

Heine, A., Thaler, V., Tamm, S., Hawelka, S., Schneider, M., Torbeyns, J., De Smedt, B, et al. (2010). What the eyes already 'know': Using eye movement measurement to tap into children's implicit numerical magnitude representations. *Infant and Child Development, 19*, 175-186.

Huber, S., Moeller, K., & Nuerk, H. (2014). Dissociating number line estimations from underlying numerical representations. *The Quarterly Journal of Experimental Psychology, 67*(5), 991-1003.

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive Science, 37*, 775-799.

Landy, D., Allen, C., & Zednik, C. (2014). A perceptual account of symbolic reasoning. *Frontiers in Psychology, 5*, 275.

Landy, D., Charlesworth, A., & Ottmar, E. (2017). Categories of large numbers in limne estimation. *Cognitive Science, 41*(2), 326-353.

Laski, E.V., & Dulaney, A. (2015). When prior knowledge interferes, inhibitory control matters for learning: The case of numerical magnitude representations. *Journal of Educational Psychology, 107*(4), 1035-1050.

LeFevre, J., Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, J.L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology, 4*, 1-9.

Levine, D. R. (1982). Strategy use and estimation ability of college students. *Journal for Research in Mathematics Education, 13*(5), 350-359.

Link, T., Nuerk, H.C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The Quarterly Journal of Experimental Psychology, 67*(8), 1597-1613.

Loftus, A.M., Nicholls, M.E., Mattingley, J.B., & Chapman, H.L. (2009). Pseudoneglect for the bisection of mental number lines. *The Quarterly Journal of Experimental Psychology, 62*(5), 925-945.

Meyerhoff, H.S., Moeller, K., Debus, K., & Nuerk, H.C. (2012). Multi-digit number processing beyond the two-digit number range: A combination of sequential and parallel processes. *Acta Psychologica, 140*, 81-90.

Mock, J., Huber, S., Klein, E., & Moeller, K. (2016). Insights into numerical cognition: considering eye-fixations in number processing and arithmetic. *Psychological Research, 80*(3), 334-359.

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

Newman, K., Xu, C., LeFevre, J. (2017, February). Strategy sophistication and accuracy: Adult performance on the number line task. Poster session presented at the Royal Society, London, UK.

Nuerk, H., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition, 82*(1), 25-33.

Nuerk, H.C., & Willmes, K. (2005). On the magnitude representations of two-digit numbers: Three models of two-digit magnitude representation. Psychology Science, 47(1), 52-72.

Nuerk, H., Moeller, K., Klein, E., Willmes, K., & Fischer, M. (2011). Extending the mental number line: A review of multi-digit number processing. *Joural of Psychology, 219*(1), 3-22.

Olsen, A.L. (2013). Leftmost-digit-bias in an enumerated public sector? An experiment on citizens' judgment of performance information. *Judgment and Decision Making, 8*(3), 365-371.

Opfer, J. (2003). Analyzing the number-line task: A tutorial. Retrieved July 19, 2017, from http://www.psy.cmu.edu/~siegler/SiegOpfer03Tut.pdf

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

Poltrock, S.E., & Schwartz, D.R. (1984). Comparative judgments of multidigit numbers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10(1), 32-45.

Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., & Luwel, K. (2016). Children's use of number line estimation strategies. *European Journal of Psychology of Education, 31*(2), 117-134.

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

Reys, R., & Bestgen, B. (1981). Teaching and assessing computational estimation skills. *Elementary School Journal, 82*(2), 116-127.

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science, 17*(4), 525-536.

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smet, B., Verschaffel, L., Jacobs, A., & Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development, 23*, 424-437.

Schneider, M., Grabner, R.H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical school achievement: their interrelations in grades 5 and 6. *Journal of Education Psychology, 101*, 359-372.

Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2015). Varieties of quantity estimation in children. *Developmental Psychology, 51*(6), 758-770.

Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2017). Preschool children use space, rather than counting, to infer the numerical magnitude of digits: Evidence for a spatial mapping principle. *Cognition, 158*, 56-67.

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

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

Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: The role of visuomotor integration and visuospatial skills. Journal of Experimental Child Psychology, 145, 22-33.

Slusser, E., & Barth, H. (2017). Intuitive proportion judgment in number-line estimation: Converging evidence from multiple tasks. *Journal of Experimental Psychology, 162*, 181-198.

Slusser, E., Santiago, R., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology, 142*(1), 193-208.

Sowinski, C., LeFevre, J., Skwarchuk, S.L., Kamawar, D., Bisanz, J., & Smith-Chant, B. (2015). Refining the quantitative pathway of the Pathways to Mathematics model. *Journal of Experimental Child Psychology 131*, 73-93.

Stapel, J. C., Hunnius, S., Bekkering, H., & Lindemann, O. (2015). The development of numerosity estimation: Evidence for a linear representation early in life. *Journal of Cognitive Psychology, 27*(4), 400-412.

Steiner, E., & Ashcraft, M. (2012). Three brief assessments of math achievement. *Behavior Research Methods, 44*(4), 1101-1107.

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

Thompson, J.M., Nuerk, H.C., Moeller, K., & Kadosh, R.C. (2013). The link between mental rotation ability and basic numerical representations. *Acta Psychologica, 144*, 324-331.

Thompson, C., & Opfer, J. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81*, 1768-1786.

van Viersen, S., Slot, E. M., Kroesbergen, E. H., van't Noordende, J. E., & Leseman, P. P. (2013). The added value of eye-tracking in diagnosing dyscalculia: A case study. *Frontiers in Psychology, 4*(679), 1-10.

van't Noordende, J. E., van Hoogmoed, A. H., Schot, W. D., & Kroesbergen, E. H. (2016). Number line estimation strategies in children with mathematical learning difficulties measured by eye tracking. *Psychological Research, 80*(3), 368-378.