SELECTION OF PROCEDURES IN MENTAL DIVISION: RELATIONS BETWEEN SELF-REPORTS AND EYE-MOVEMENT PATTERNS

by

Matthew G. Huebner

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

Master of Arts in

Psychology

Carleton University Ottawa, Canada

Abstract

Do eye-movement patterns reflect the procedures people use when solving basic arithmetic problems? Sixty-eight adults solved simple division problems while their eye movements were recorded. Thirty-four of these participants reported their solution processes (Experiment 1A: Self-Report Condition) and 34 participants did not (Experiment 1B: Combined Analyses). Participants in Experiment 1A were classified into procedure groups based on their reported use of procedures for large division problems: Retrievers, transformers, and counters. Transformers and counters fixated more on the left and right operands than retrievers for large problems. The 34 participants in Experiment 1B were categorized based on their values of mu and tau for large problems. Patterns of performance for these participants, combined with those who provided self-reports, complemented the patterns found in Experiment 1A. The above results lend support to the use of eye tracking to augment traditional measures of performance when assessing individual differences in procedure selection.

Table of Contents

Acknowledgements

My personal thanks goes first and foremost to my amazing supervisor, Dr. Jo-Anne LeFevre. Her constant care, support, and trust allowed me to pursue my own research interests and this thesis is a better product as a result.

I'd also like to thank the most incredible committee a graduate student could have asked for, including Dr. Craig Leth-Steensen, Dr. Chris Herdman, Dr. Jim Davies, and Dr. Tina Daniels. Your feedback and suggestions were invaluable.

Thanks to Richard Vachon and Graeme Schmidt, two amazing individuals I'm lucky enough to call close friends. You two were always willing to listen to my struggles and reminded me to have fun once in a while.

Finally, I want to thank my mom and dad for all their love and support over the years.

List of Tables

List of Figures

Selection of Procedures in Mental Division: Relations between Self-Reports and Eye-Movement Patterns

Skilled adults, like children, demonstrate a remarkable degree of procedural flexibility when they solve basic arithmetic problems (LeFevre, Sadesky & Bisanz, 1996; LeFevre, Bisanz et al., 1996; Campbell & Xue, 2001; Robinson, Arbuthnott, & Gibbons, 2002). This procedural flexibility is one explanation for the problem-size effect (Groen & Parkman, 1972; Zbrodoff & Logan, 2005), where solvers become slower and less accurate as problems become larger (e.g., 9 / 3 → 72 / 8). Adults appear to use fast and efficient memory retrieval on small problems versus slow, less-efficient mental algorithms on large problems. To date, the most common method of assessing non-retrieval procedure use in mental arithmetic is self-reports, where participants describe their solution processes after each problem (Campbell & Austin, 2002; LeFevre et al. 1996; Robinson et al., 2002). However, the use of introspective methods has been challenged. Researchers argue that participants are either reactive to task demands (Kirk & Ashcraft, 2001), or cannot consciously detect their solution processes and thus self-reports are not veridical (Fayol & Thevenot, 2012). Accordingly, developing alternative approaches for assessing procedure use is important for contemporary research in cognitive arithmetic. The goal of this thesis was to examine procedure use in basic division (e.g., 27 / 3) using (1) a distributional analysis of reaction times and (2) eye-movement patterns in the form of gaze duration and fixation counts across problems.

Non-Retrieval Procedures in Mental Arithmetic

In historical models of the problem-size effect, adults were assumed to solve all basic arithmetic problems via direct memory retrieval (e.g., Ashcraft, 1987; Groen &

Parkman, 1972). Children, in contrast, were shown to use non-retrieval procedures in the process of acquiring basic arithmetic knowledge (Groen & Parkman, 1972; Siegler, 1989). As such, the problem-size effect in adults was attributed primarily to the way in which arithmetic facts are stored in memory (McCloskey, Harley, & Sokol, 1991). Ashcraft and Battaglia (1978) proposed a structural account of arithmetic fact retrieval, whereby arithmetic facts are stored in a mental table. Rows represent one operand and columns represent another operand. The bigger the operands, the longer the search process across rows and columns to find the corresponding sum. For example, finding the sum to the question 2 + 3 takes less time than 7 + 8 because 2 + 3 only requires a single row shift downward and two column shifts to the right.

In contrast to structural models of the problem-size effect, other researchers promoted the idea of an associative network. For instance, Ashcraft (1987) proposed a network retrieval model of arithmetic fact retrieval. In this model, two distinct sets of nodes represent the operands for a given problem. For example, the problem 5 + 4 would activate the node for 5 in the first set and activate 4 in the second set. Each operand node is connected to various potential solutions, in that 5 is associated with 6, 7, 8, 9, 10, etc. Likewise, the operand 4 is connected with potential solutions that involve that operand (5, 6, 7, 8, 9, etc). The strength of association between operand nodes and solution nodes is proportional to the frequency of presentation of each problem. Because problems with small operands are generally encountered more frequently (Ashcraft & Christy, 1995), associations between small operands and potential solutions are stronger than associations between large operands and potential solutions. Thus, small problems are solved faster than large problems.

However, a primary issue with these models was their inability to explain the patterns of results for tie problems (e.g., 3 + 3, 6 x 6). Adults frequently solve tie problems faster than similarly sized non-tie problems (e.g., 6 + 6 vs. 6 + 7) and ties do not show problem-size effects (LeFevre, Sadesky, & Bisanz, 1996; LeFevre, Bisanz, et al., 1996). This result led to the suggestion that tie problems have special representations in memory. LeFevre, Sadesky, and Bisanz (1996) proposed that adults are flexible in the types of procedures they use to solve basic arithmetic problems. In their study, participants solved basic addition problems with operands ranging between 0 and 9. Participants solved each problem under instructions to respond quickly and accurately and also described how they solved the problem after each trial. Although memory retrieval was reported most, these adults also reported using non-retrieval procedures. These non-retrieval procedures included counting (e.g., 5 + 3 = 6 → 7 → 8), rules (anything plus 0 is itself) and decomposition (e.g., 7 + 8 = 7 + 7 + 1 = 15). A series of linear regressions for latencies showed that problem size was a better predictor of latencies when zero and tie problems were removed from the data set. Furthermore, latencies for trials where participants used a counting procedure accounted for substantially more variance in the problem-size effect than trials where retrieval was reported. Furthermore, participants showed an increase in the rate of transformations and counts as the sum of the problem increased.

These results were also extended to multiplication performance. Across two experiments, LeFevre, Bisanz et al. (1996) had participants solve single-digit multiplication problems with operands ranging from 0 to 9. As with the addition experiment, participants reported their solution procedures after each trial. Participants used memory retrieval more frequently for multiplication than addition, but still reported various non-retrieval procedures to solve some problems. Such procedures included the zero rule (anything times zero is zero), decomposition (e.g., 8 x 9 = [8 x 8] + 8) number series (e.g., 4 x 5 = 5, 10, 15, 20) and repeated addition (e.g., 4 x 3 = 4 + 4 + 4). Based on these findings, the authors concluded that adults, much like children, use a variety of non-retrieval procedures to solve basic arithmetic problems.

Further evidence that adults use non-retrieval procedures comes from studies that investigate the relationship between two operations. Researchers have shown that adults rely on their knowledge of multiplication facts to solve division problems (LeFevre & Morris, 1999; Mauro, LeFevre, & Morris, 2003; Robinson et al., 2002) and also rely on their knowledge of addition facts to solve subtraction problems (Campbell, 2008). For example, LeFevre and Morris (1999) had participants solve simple multiplication and division problems under two experimental conditions. In the first condition, participants solved multiplication problems, followed by division problems. In the second condition, participants solved division problems, followed by multiplication problems. LeFevre and Morris (1999) hypothesized that if participants rely on their knowledge of multiplication to solve division problems, then performance on multiplication problems when they were presented in the second block of trials should be better than when they were presented in the first block. As predicted, participants who solved multiplication problems after division were faster and more accurate than participants who solved multiplication problems before division. LeFevre and Morris (1999) concluded that adults engage in a mediated retrieval procedure, in which they recast division problems as multiplication problems to accurately reach a solution. Thus, solving division problems provided multiplication "practice."

Additional evidence for the mediated retrieval hypothesis comes from Mauro et al. (2003), who manipulated the surface format of multiplication and division problems to influence patterns of performance. In particular, division and multiplication problems were rearranged into their respective transformed formats (e.g., 56 / 8 → 8 x __ = 56; 42 / 6 → __ / 6 = 7). Mauro et al. (2003) predicted that if participants rely on their knowledge of multiplication to solve division problems, participants would show a latency and accuracy advantage for division problems already presented in the transformed format. Consistent with this prediction, participants who solved division problems in the transformed format showed both a latency and accuracy advantage over problems solved in the standard format. In addition, multiplication problems in the transformed division format showed increased latencies and errors. Critically, the observed advantage for transformed division problems only occurred when problem size was large (dividend greater than 25). Based on this finding, adults may rely on their knowledge of multiplication to aid in solving division problems, especially as problem size increases.

Studies investigating mediated retrieval also extend to the relationship between addition and subtraction. Campbell (2008) noted that, based on participants' self-reports, people tend to solve subtraction problems by referencing the corresponding addition problem. For example, when participants are presented with a subtraction problem such as 17 – 9, they may transform that problem into the equivalent addition format 9 + ? = 17 to find the solution. Following the experimental manipulations of Mauro et al. (2003), participants were asked to solve basic subtraction problems presented either in a standard or transformed format. Participants solved large subtraction problems faster if they were presented in the addition format than if they were presented in subtraction format. In contrast, participants solved small subtraction problems faster when they were presented in the standard format. Campbell (2008) concluded that the performance advantage for large subtraction problems in the transformed format reflected a tendency for adults to rely on their knowledge of addition when solving large subtraction problems. The performance advantage for small subtraction problems in the standard format was assumed to reflect use of memory retrieval.

Further support for the hypothesis that adults use non-retrieval procedures comes from cross-cultural studies. For example, Campbell and Xue (2001) compared performance of Canadian Chinese, Asian Chinese, and Non-Asian Canadian students who solved basic arithmetic problems across four operations. Non-Asian Canadian participants elicited the greatest problem-size effect across all four operations and also reported the largest number of non-retrieval procedures. In contrast, there were no significant differences in performance between Canadian Chinese participants and Asian Chinese participants. Based on this finding, Campbell and Xue (2001) suggested that differences in performance across cultural groups are not based solely on variation in formal schooling. Rather, performance differences across cultures are the result of culture-specific factors, such as how much one culture values mathematical knowledge and fluency.

In summary, traditional models of arithmetic fact retrieval ignored the possibility that adults may use non-retrieval procedures and thus did not address how non-retrieval procedures influence adults' performance in simple arithmetic tasks. Contrary to the view that answers are always retrieved from memory, adult participants show a substantial amount of flexibility in their use of non-retrieval procedures when solving basic arithmetic problems. Evidence for this procedural flexibility generally stems from the use of self-reports that complement traditional dependent measures of performance. Adults typically rely on their knowledge of one operation to solve problems involving a different operation and even show variability in procedure selection based on cultural factors such as amount of practice. The question of whether self-reports are veridical measures of cognitive processing, however, remains a topic of considerable debate.

Challenges to the Self-Report Methodology

Although many researchers agree that self-reports have provided substantive insight into the types of procedures adults use when solving basic arithmetic, others remain skeptical. Some researchers claim that participants are highly reactive to explicit task demands (Kirk & Ashcraft, 2001), whereas others claim that the observed patterns of procedure reports are sensitive to indirect task manipulations (Campbell & Austin, 2002). Furthermore, some researchers postulate that non-retrieval procedures sometimes occur too quickly to be consciously detected (Fayol & Thevenot, 2012; Barrouillet & Thevenot, 2013).

Kirk and Ashcraft (2001) hypothesized that individuals are reactive to the explicit demands of a particular task and may alter the pattern of procedures they report when solving simple arithmetic problems. Adults solved problems in one of four experimental conditions: a silent control condition, a self-report condition, and two biasing conditions. One biasing condition emphasized the use of memory retrieval and the other emphasized the use of non-retrieval procedures. In their first experiment, Kirk and Ashcraft (2001) found that adults in the retrieval-biased condition reported using retrieval on 90% of addition trials. In contrast, participants in the procedure-biased condition reported using retrieval on only 32% of trials. In their second experiment, participants in the retrieval-biased condition reported retrieval for 96% of multiplication trials, whereas participants in the procedure-biased condition reported retrieval for 62% of trials. The authors concluded that adults are reactive to explicit task demands and cautioned future researchers about using self-reports to draw conclusions about cognitive processes.

The impact of task demand and instructional bias also appears to vary across participants with various skill levels. For instance, Smith-Chant and LeFevre (2003) classified adults as highly-, average-, or low-skilled in arithmetic based on their performance on a multi-digit arithmetic task. Participants then solved single-digit multiplication problems under different instructional conditions that emphasized either speed or accuracy. Some conditions also required participants to report how they solved the problem after each trial. Low-skilled participants reported using more non-retrieval procedures than average- or highly-skilled participants. In the self-report condition, low-skilled participants reported using non-retrieval procedures more often when accuracy was emphasized over speed. Thus, the combination of participant skill and instructional demand appears to influence the patterns of performance when adults solve basic arithmetic.

Not only do self-reports appear to be influenced by explicit task demands, but they also appear to be influenced other experimental manipulations. For instance, Campbell and Austin (2002) had adults solve simple addition problems under speeded and delayed response deadlines. On large problems, adults reported using direct memory retrieval more often in the speeded response condition (solve the problem before the onset of a beep at 750 ms) than in the delayed response condition (wait for 2500 ms after stimulus presentation). Campbell and Austin (2002) concluded that manipulating response deadlines affect the frequency in which adults use retrieval when solving basic arithmetic.

Adults also may simply be unaware of the types of procedures they use when solving basic arithmetic problems. For example, Fayol and Thevenot (2012) presented participants with simple addition, subtraction, and multiplication problems under various stimulus presentation conditions (Experiment 1). There were three conditions: Problems (a) appeared in their entirety (e.g., 2 + 3), (b) the operator appeared 150 ms before the operands (e.g., + → 2 + 3), or (c) operands appeared 150 ms before the operator (e.g., 2 3 → 2 + 3). Participants were faster at solving addition and subtraction problems when the operator appeared before the operands themselves, but no similar RT advantage occurred for multiplication problems.

Fayol and Thevenot (2012) suggested that multiplication is usually solved via rote memory retrieval, and therefore no RT advantage should be observed for multiplication problems when the operator is presented first. In contrast, participants generally rely more on non-retrieval procedures for addition and subtraction. As such, the presentation of the operator before the operands may have primed an abstract non-retrieval procedure that caused an RT advantage to occur. In a second experiment, the results for addition and multiplication problems were replicated, this time with small addition problems. Contrary to previous research that suggested almost exclusive use of retrieval for small addition problems, Fayol and Thevenot (2012) proposed that abstract, non-retrieval procedures occur just as frequently for small problems. As a result, the use of self-reports to assess procedure use on small arithmetic problems may not reflect the actual cognitive processes responsible for performance.

Further support for the idea that adults use abstract procedures at an unconscious level comes from Barrouillet and Thevenot (2013). In this study the authors presented participants with small addition problems (i.e., problems with a sum no larger than 8) and found a small linear increase in RT as a function of left and right operand size. The authors also found a left operand x right operand interaction that was primarily accounted for by the inclusion of tie problems. The authors attributed these findings to the use of a fast and efficient counting procedure, as opposed to a decrease in retrieval efficiency. Specifically, if the problem-size effect is observable even for very small arithmetic problems, participants possibly use non-retrieval procedures irrespective of problem difficulty.

In summary, traditional introspective methods of assessing cognitive processes continue to generate controversy. As in many experimental contexts, participants display a relatively high degree of reactivity to task demands. Emphasizing a particular approach may influence reported solution procedures. Procedure use also appears to be affected by task instructions (e.g., speed vs. accuracy), as well as participant skill level and by response delays. Finally, participants may use efficient, non-retrieval procedures that fall below the threshold of conscious awareness. As of result of these studies, researchers have relied on alternative, more objective approaches to assess procedure use in skilled adults. One such approach is the use of an ex-Gaussian analysis to examine changes in parameter values across small and large problems.

Seeking Objectivity: Assessment of Procedure Use through Decomposition of the Mean

As a response to challenges regarding the self-report methodology, researchers have sought alternative approaches to assessing the use of non-retrieval procedures in adults' arithmetic performance. Based on previous literature that used introspective methods, retrieval of a solution from memory is typically fast and efficient. Conversely, non-retrieval procedures are more time-consuming, as they require a series of mental steps or manipulations to generate the correct solution. Therefore, the RT distributions of non-retrieval trials should have slower response latencies and more positive skew than the RT distributions of retrieval trials. However, many researchers fail to consider the skew of RT distributions or to relate the distributions to different solution procedures.

Landmark studies that examined non-retrieval procedure use in adults either analyzed median RTs (LeFevre, Sadesky, & Bisanz, 1996; LeFevre, Bisanz, et al., 1996) or trimmed 'extreme' latencies to induce normality before data analysis (Campbell & Timm, 2000; Campbell & Xue, 2001). Although trimming is a common practice when dealing with RTs, many researchers discard extreme scores under the false assumption that such scores merely represent a nuisance event during the experiment (such as the participant spacing out or blinking during stimulus presentation). According to Heathcote, Popiel, and Mewhort (1991), RTs falling within the tail of a distribution may represent a specific cognitive process of interest. Thus, researchers who discard or ignore response times that do not fall within the normal curve may be ignoring the very process under empirical investigation.

A three-parameter model known as the ex-Gaussian can characterize skewed distributions and thus may be a better fit to RT distributions than symmetrical, Gaussian models. Mathematically speaking, the ex-Gaussian is a convolution of the normal and exponential distributions. The leading edge or peak of the distribution, known as mu, represents the area under the curve where the majority of scores fall. Tau represents the mean of the exponential component that makes up the distribution, such that larger tau values indicate greater skew. These parameters are related such that mu + tau = the mean. Researchers who have used ex-Gaussian analyses have obtained insights into the underlying cognitive processes for a variety of psychological phenomena that extend or augment those available from traditional analyses (Heathcote et al., 1991; Hockley, 1984; Leth-Steensen, Elbaz, & Douglas, 2000; Ratcliff & Murdock, 1976, Ratcliff, 1979).

Penner-Wilger, Leth-Steensen, and LeFevre (2002) postulated the ex-Gaussian model would provide useful insight into the procedures adults use when solving basic arithmetic problems. In particular, they proposed that different solution procedures for single digit multiplication problems could be represented by differences in mu and tau values across problem size. Specifically, the value of the normal component was assumed to represent the efficiency by which participants retrieve answers from memory. Therefore, an increase in the normal component represents a decrease in retrieval efficiency (i.e., participants take longer to retrieve an answer directly from memory). In contrast, values in the tail were assumed to represent slow, non-retrieval procedures, and thus an increase in the mean of the tail represents an increased reliance on non-retrieval procedures. This characterization makes sense if non-retrieval procedures are more time-consuming than direct retrieval.

To test this interpretation of the ex-Gaussian model as applied to arithmetic, Penner-Wilger et al. (2002) compared data collected from Canadian- and Chinese-educated adults who performed a simple multiplication task. Because Canadian adults report using non-retrieval procedures whereas Chinese adults do not (Campbell & Xue, 2001), it was predicted that Canadian participants would elicit larger tau values as a function of problem size. In accordance with this prediction, Chinese participants showed only small differences in mu for small versus large problems, whereas Canadian subjects showed larger mu and tau values across problem size. These patterns were consistent with the assumption that Chinese adults use retrieval almost exclusively on all problems, with a slight decrease in retrieval efficiency on large problems as reflected by a shift in mu. However, Canadian adults relied more on time-consuming non-retrieval procedures as problems became larger. This increase in non-retrieval procedure use is reflected by an increase in the mean of the tail as problem size increases.

One limitation of Penner-Wilger et al.'s analysis was that trial-by-trial self-reports of procedure use were not obtained in the experiment. Information about procedure use was gathered during interviews after data collection was completed (LeFevre & Liu, 1997). Using self-reports on a trial-by-trial basis, Campbell and Penner-Wilger (2006) assessed changes in procedure use when addition problems varied in size (small vs. large) and presentation format (digit vs. word). The authors hypothesized that problems presented in word format would disrupt retrieval, resulting in a format x size interaction. Consistent with the results reported by Penner-Wilger et al. (2002), mu and tau values differed as a function of problem size. As problem size increased, participants reported using more non-retrieval procedures and, as a result, elicited larger tau values.

Interestingly, for the large problem/word format condition, non-retrieval procedure trials were reflected in mu and not tau. Specifically, retrieval trials and procedure trials both contributed equally to mu, as the large problem/word format condition produced substantially more non-retrieval procedures. Campbell and Penner-Wilger (2006) concluded that non-retrieval procedures are not merely reflected by a single parameter of RT distributions.

Variations in mu and tau values have also been used to validate self-reports and understand individual differences in procedure use for simple subtraction. For instance, LeFevre et al. (2006) used self-reports to categorize participants into specific groups, with mu and tau values calculated for each participant on small and large problems (e.g., 6 – 2 vs. 17 – 5). Participants who reported using retrieval on the majority of large problems were placed into the *retriever* group, whereas participants who frequently relied on transformations or counting were placed into the *transformer* and *counter* groups, respectively. Retrievers only showed a minor increase in tau with increased problem size, with a negligible shift in mu. Transformers showed shifts in mu and increases in tau. Counters showed the largest shifts in mu and increases in tau. LeFevre et al. (2006) concluded that both parameters offer insight into the extent to which participants use retrieval and procedural strategies. Participants who primarily use retrieval maintain efficiency across problem size, and therefore do not produce a shift in mu. Participants who use transformations to solve the majority of large problems had larger values of mu, because transformations are less efficient than retrieval. Finally, participants who relied on counting for large problems show significant differences in both parameters across problem size. In this case, use of retrieval on small problems versus transformation on large problems is reflected by a shift in mu, with time-consuming counting trials built up in the tail of the distribution. Thus, tau might reflect the extent to which individuals use a counting procedure for large problems.

LeFevre et al. (2006) also showed that mu and tau values could be used to group participants into procedure groups even without self-reports. Mu and tau values were calculated for an additional group of participants who were not required to report their solution procedures. These values were then combined with those of the participants in the self-report condition in a discriminant function analysis. This analysis correctly classified 69% of the original participants as retrievers, transformers, or counters (originally categorized with self-reports) using patterns of mu and tau for large problems. Patterns of performance for mu and tau values across groups based on the discriminant function analysis appeared similar to the patterns based on participants' self-reports alone. This result suggested that mu and tau could accurately differentiate between procedure groups even when self-reports are not used.

In summary, response time distributions generally reflect multiple cognitive processes that influence a particular psychological phenomenon. As such, assuming normality may result in a loss of important data that represents a cognitive process of empirical interest if the distribution is not normal. In the case of mental arithmetic performance, non-retrieval procedures are slower and therefore reflect the build-up in the skew of the distribution. However, changes in both mu and tau values should be considered when assessing patterns of performance. Finally, mu and tau values appear to be useful variables for assessing individual differences even when self-reports are not available.

A Second Alternative: Procedure Use Reflected by Eye Patterns

Only a few published studies have used eye movements to examine arithmetic performance. For example, Verschaffel, De Corte, Gielen, and Strupf (1994) examined fixation patterns made by children when they solved addition problems involving three numbers (e.g., 5 + 7 + 4). Children's eye movements were recorded while they solved each problem. In conjunction with eye movements, children also reported which numbers they began to add together after each trial. Verschaffel et al. (1994) found that children's fixation patterns complemented their verbal descriptions of how they solved each problem. Children would typically report rearranging the numbers in that two identical digits could be added together first (e.g., 2 + 5 + 2) or that the first two digits would equal ten.

Green, Lemaire, and Dufau (2007) investigated whether fixation patterns complemented procedure use when adults solved multi-digit addition problems (e.g., 342 + 479). Two procedures are typically used when multi-digit arithmetic problems are solved mentally: First, participants start adding by the units, followed by the tens, and then the hundreds. The second procedure operates in the reverse fashion, where participants start by adding the hundreds. Participants were randomly assigned to choice and no-choice procedure conditions. In the choice condition, participants were allowed to use either procedure. In the no-choice condition, participants were told which procedure to use for a particular problem. In the choice condition, fixation patterns complemented participants' self-reports. For instance, fixations began on the left side of the problem when participants reported using the hundreds procedure. Likewise, fixations began on the right side of the problem when participants reported using the unit procedure. In the no-choice condition, participants' fixation patterns also complemented the particular procedure they were informed to use.

The use of eye movements has also complemented specific hypotheses related to mental representations of arithmetic facts. Zhou, Zhao, Chen, and Zhou (2012) investigated eye movements of post-secondary students in China while they solved single-digit multiplication and addition problems. The authors noted that the Chinese multiplication table contains only a single representation for each possible pair of digits, each beginning with the smaller operand. For example, the table would contain 7 X 9 but not 9 X 7. The authors predicted that Chinese students should show a preference in operand order. As predicted, participants made rightward eye movements when the problem began with the smaller operand and leftward movements when the problem began with the larger operand. Zhou et al. (2012) concluded that eye movements plausibly reflect how stored arithmetic facts are accessed and manipulated.

Curtis, Huebner, and LeFevre (2014) examined changes in gaze duration and fixation count patterns across all four basic operations (i.e., addition, subtraction, multiplication, and division). Their primary goal was to examine how eye patterns change across problem size, depending on the operation being tested. Each problem was divided into three areas of interest: one for each operand and one for the operator. Main effects of problem size and interest area were found for all four operations. Specifically, participants looked at large problems longer than they looked at small problems. Furthermore, total gaze duration was not divided evenly across interest areas. For addition and multiplication, participants allocated most of their attention to the operator, irrespective of problem size. In contrast, participants showed a robust problem size by interest area interaction for subtraction and division. For these two operations, gaze durations and number of fixations were fairly even across interest areas when problem size was small. However, participants began placing a greater emphasis on the operands as problems became larger. Specifically, participants looked more at the first operand for large division problems and more at the second operand for large subtraction problems. To date, this is the first study to examine how the problem-size effect is distributed across problem components.

Curtis et al. (2014) postulated that the robust problem size by interest area interaction observed for subtraction and division may reflect a tendency for participants to switch from retrieval to non-retrieval procedures as problem size increases. This approach makes sense considering the available literature on the relationship between operations (LeFevre & Morris, 1999; Mauro et al., 2003; Campbell, 2008). However, Curtis et al. (2014) did not collect self-reports and thus we were unable to examine whether non-retrieval procedures contributed to the observed differences in eye movement patterns across problem size. Furthermore, we were unable to explore individual differences in the selection of procedures when participants solved subtraction and division problems.

To summarize, eye movements have rarely been used to examine performance in simple arithmetic tasks. Available research appears promising in that specific procedures are reflected through specific patterns of fixations for complex arithmetic problems and may also reflect how facts are stored in memory. Patterns of gaze and fixations appear to differ across operations for simple arithmetic problems, especially for operations that promote greater use of non-retrieval procedures as a function of problem size. Although the available research is limited at the present time, the increased attention paid to the operands for large subtraction and division problems (relative to the operator) in simple arithmetic tasks may reflect a shift from retrieval to non-retrieval procedures and provide greater insight into how these procedures are utilized.

The Present Research

Although some studies have related the shape of RT distributions to specific solution procedures, the use of eye movements to assess procedure use has been far less common (Zhou et al., 2012). Research on procedure selection has consistently demonstrated that participants report greater use of non-retrieval procedures on division and subtraction than on addition and multiplication, especially for large problems (Campbell & Xue, 2001; Robinson et al., 2002). In accord with this finding, gaze duration and fixation count patterns appear to deviate substantially as problem size increases for division and subtraction problems (Curtis et al., 2014). Thus, the goal of this thesis was to assess how eye movement patterns are affected when adults of various skill levels (i.e., participants who primarily use memory retrieval vs. participants who readily use non-retrieval procedures) solve simple division problems.

Considering the above findings, a number of key hypotheses and predictions could be made. First, if parameters of the ex-Gaussian distribution accurately complement self-reports, then participants who frequently report the use of non-retrieval procedures such as recasting and counting should elicit systematic patterns in mu and tau that logically complement their choices of procedures. Specifically, transformers and counters should elicit larger increases in tau when compared to retrievers since repeated subtraction and relying on derived facts (i.e., multiplication) in division is considered slower than direct retrieval. Second, if the increased attention paid to the operands for large division problems reflects non-retrieval procedures, then participants who rely more on recasting and counting should show a greater problem size x interest area interaction than participants who primarily rely on direct memory retrieval. Finally, if mu and tau values for large division problems are indicative of individual differences in the selection of non-retrieval procedures, then procedure groups based on these values should elicit similar eye-movement patterns across division problems, even when self-reports are not required.

Method

The research described was broken down into Experiment 1A and Experiment 1B. Experiment 1A (i.e., the Self-Report condition) was designed to address two primary goals. First, I wanted to replicate patterns of performance for division that were reported by Curtis et al. (2014). Replicating these findings would support the claim that overall performance is not drastically affected when participants provide self-reports. The second goal was to categorize participants into procedure groups based on their reported use of memory retrieval, recasting, and counting for large division problems. This categorization process would allow for comparisons of performance across procedure groups and assess differences in eye-movement patterns. The results of Experiment 1A are based on the same data presented by Huebner, MacKay, and LeFevre (2013), with the exception that a three-group solution was utilized here. Subsequently, Experiment 1B included a combined analysis of the same participants in Experiment 1A and an additional sample of participants who did not provide self-reports. Here, a discriminant function analysis based on participants' mu and tau values for large division problems was used to determine procedure groups and assess whether eye-movement patterns would be similar to those found in Experiment 1A.

Participants

A total of 68 Carleton University students (32 females) participated in this experiment. They were recruited either through the online research participation pool or through recruitment posters.

Experiment 1A. Thirty-four Carleton University students (17 females). Participants had a median age of 22 years (*SD* = 4.17 years) and all reported normal or corrected-to-normal vision. Undergraduate students (*n* = 23) received 1.5% course credit and graduate students (*n* = 11) were paid $10.00.

Experiment 1B. Thirty-four Carleton University students (15 females). Participants had a median age of 21 years (*SD* = 4.41 years) and all reported normal or corrected-to-normal vision. Undergraduate students (*n* = 31) received 1.5% course credit and graduate students (*n* = 3) were paid $10.00.

Materials and Procedure

Participants were tested individually in a quiet room by a trained experimenter. The experiment lasted no longer than one hour, after which participants were debriefed and dismissed.

Arithmetic Fluency. Participants were each given the Calculation Fluency Test (Sowinski, Dunbar, & LeFevre, 2014). This test assesses individuals' level of competence with addition, subtraction and multiplication. Specifically, there were two pages for each operation tested, with 60 questions per page. Addition problems and subtraction problems were constructed using two two-digit operands (e.g., 34 + 89, 67 - 21). Multiplication problems were constructed using a two-digit multiplicand and a single-digit multiplier (e.g., 54 x 3). Each subject was given one minute per page to answer as many questions as possible while maintaining accuracy.

Strategies. The list of available procedures given to each participant was nearly identical to that used by Campbell and Xue (2001):

1) Memory: You solve the problem by just remembering or knowing the answer from memory. For example, you simply remember that 3 is the answer to the problem 15 / 5. The answer seems to jump into your head. In other words, no mental calculations or manipulations were necessary.
2) Recast: You solve the problem by referring to the related operation (in this case, multiplication) to solve the problem. For example, you are given the problem 56 / 7 and convert it to 7 x ? = 56 in order to get 8.
3) Count: You solve the problem by counting a certain number of times to get the answer. Specifically, you might use the divisor to count backwards in order to obtain the correct answer. For example, 24 / 6 = 24 – 6 – 6 – 6 – 6 = 0. Therefore, the answer is 4 because you subtracted 6 four times.
4) Other: You solve the problem by using a strategy not listed above, or you do not know the answer to the problem (e.g., you guessed).

Stimuli. Division problems were constructed by inverting all pairwise combinations of multiplication problems with operands between 2 and 9, including tie problems (e.g., 16 / 4). Commutative problems (e.g., 18 / 6 and 18 / 3) were considered different problems, resulting in 64 different problems. Problems with a dividend of 25 or less were considered small problems (e.g., 20 / 5), and problems with a dividend greater than 25 (e.g., 72 / 8) were considered large problems. Each problem was presented 3 times, resulting in 192 trials.

Procedure. Stimuli were presented visually on a 15-inch Dell plug-and-play monitor that was attached to a Dell Precision PWS390 computer, running Windows XP Professional version 2002 and EyeLink 1000 experiment software (SR Research). Participants responded verbally into a headset microphone connected to an ASIO voice trigger accurate to +/- 1 ms. An experimenter recorded responses on a standard keyboard. Eye movements were recorded using a specialized camera that emitted infrared light and recorded the reflection from the participants' corneas.

Participants were tested individually, and an experimenter was in the same room at all times. Participants first read and signed an informed consent form before beginning the experiment. Once signed, participants completed the arithmetic fluency test. The eye-tracking camera, chin rest, and chair were then adjusted. Participants sat with their heads 60 cm from the display monitor. Instructions appeared on the computer screen indicating to the participants that they would be solving simple division problems verbally while their eye movements were recorded. These instructions also emphasized that no particular strategy was better than the other. Participants were instructed to provide a solution as quickly and as accurately as possible upon presentation of each stimulus, as well as a name of the strategy used for that particular trial. If the participant was unsure of the strategy used on a specific trial, he or she was prompted to give an open, detailed description of the mental steps involved. After completing 10 practice problems the eye-tracking camera was calibrated. In the calibration, dots appeared in distal locations on the screen, and participants had to fixate on them. Once a fixation was recorded, the next dot appeared. Once the experimenter determined that the fixations had been recorded accurately, the experimental trials begun. The camera was recalibrated every 32 trials to account for minor shifts in the participant's posture.

Each trial began with a fixation point in the shape of a black rectangle, with a length of 1.5 cm and width of 1 cm. The fixation point randomly appeared in one of the corners of the screen and the problem appeared once the camera recorded a fixation in that location (minimum gaze duration of 50 ms). This procedure was utilized to ensure that participants would not be looking at the central interest area upon initial stimulus presentation. The problems were presented in black Ariel font (size: 60) on a white background and remained on the screen until the microphone recorded a response, or until a timeout period of 10 seconds. Each problem was divided into three interest areas, one for each operand and one for the operator. Each interest area had dimensions of 7 cm by 7 cm. Participants were debriefed after the last block of trials and dismissed shortly thereafter.

Results

Experiment 1A

Participants solved a total of 6,528 trials. Evaluation of the raw data revealed 458 (7%) trials that were spoiled due to equipment failure (i.e., failing to detect an utterance or firing prematurely) and a further 517 (8%) trials in which participants responded incorrectly. The remaining 5,530 valid trials were used to examine whether overall fixation performance matched the results reported by Curtis et al. (2014). Bonferroni post hoc tests were used to compare group performance and 95% confidence intervals (*CI*s) were used for the interpretation of interactions (Jarmasz & Hollands, 2009).

Participant behavior was compared to the results reported by Curtis et al. (2014) to ensure the implementation of self-reports did not affect patterns of performance. Consistent with previously reported findings, participants were slower on large problems than on small problems, $t(33) = 6.76$, $p < .001$, $\eta_p^2 = .58$. Participants were also less accurate on large problems than on small problems, $t(33) = 5.19$, $p < .001$, $\eta_p^2 = .45$.

Table 1

Experiment 1A: Summary of separate 2(problem size: small, large) by 3 (interest area: left operand, operator, right operand) ANOVAs for overall patterns of gaze and number of fixations.

		Dependent Variables			
		Gaze Duration		Fixation Count	
Independent Variables	*df*	*F*	η_p^2	*F*	η_p^2
Problem Size	1,33	38.79***	.54	60.80***	.65
Interest Area	2,66	4.78*	.13	6.61**	.17
PS x IA	2,66	43.04***	.57	60.72***	.65

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Mean gaze durations and number of fixations were analyzed in separate 2 (Problem Size [small, large]) x 3 (Interest Area [left operand, center, right operand]) repeated-measures analyses of variance (ANOVA) to determine whether performance patterns were similar to those reported by Curtis et al. (2014). Results from these ANOVAs are displayed in Table 1. For gaze duration, participants looked at large problems longer than small problems (439 vs. 294 ms). Participants' looking times varied with interest area (left operand = 469 ms; operator = 331 ms; right operand = 298 ms), with a significant difference between the left and right operands ($p < .01$). There was also a problem size x interest area interaction. As shown in Figure 1, participants spent more time looking at the operator than at the operands for small problems, but looked more at the operands than at the operator for large problems.

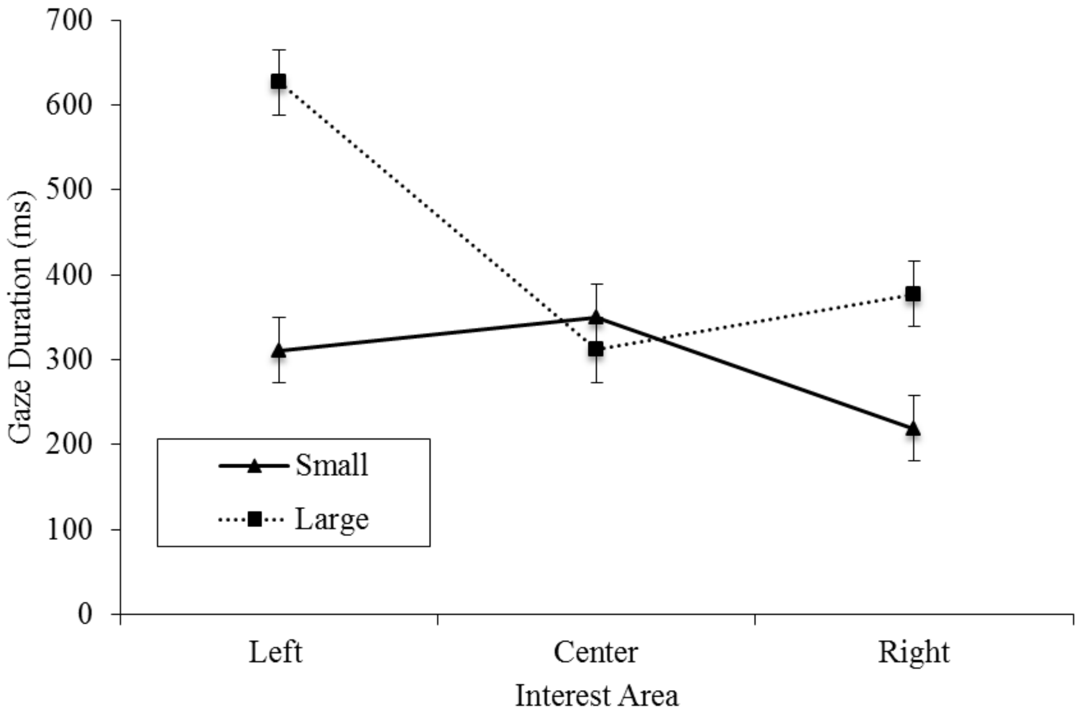

Figure 1. Experiment 1A: Mean gaze durations (ms) across interest areas for small and large problems.

The results for number of fixations were very similar. Participants made more fixations on large problems than on small problems (1.61 vs. 1.18), $F(1, 33) = 60.80$, $p < .001$, $\eta_p^2 = .65$. Participants also made an unequal amount of fixations across interest areas (left operand = 1.69; operator = 1.42; right operand = 1.08), with a significant difference between the left and right operands ($p < .001$). There was also a problem size x interest area interaction. As shown in Figure 2, participants made most of their fixations on the operator for small problems, but made most of their fixations on the operands for large problems. Thus, both gaze durations and mean fixations showed characteristic patterns by interest area, with small problems showing an inverted-v shape whereas large problems showed a typical v-shaped pattern. Curtis et al. (2014) reported nearly identical patterns, suggesting that the inclusion of self-reports had a minimal impact on performance.

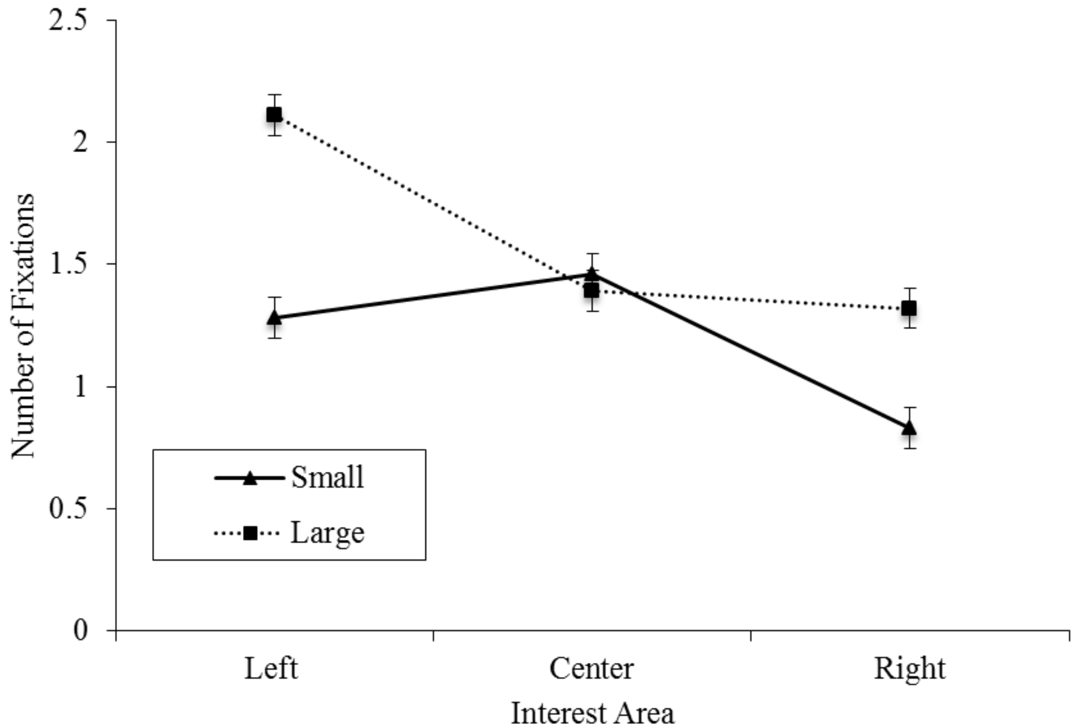

Figure 2. Experiment 1A: Mean number of fixations across interest areas for small and large problems.

Group Categorization

Participants varied considerably in the types of procedures they reported using for large division problems. Although some participants reported exclusive use of memory retrieval, others relied entirely on mental procedures such as recasting to find the correct solution. Memory was reported the most (*M* = 53%, *SD* = 25%, range of 0 to 100%), followed by recasting (*M* = 40%, *SD* = 27%, range of 0 to 100%) and counting (*M* =

11%, *SD* = 20%, range of 0 to 81%). Further examination of self-reports for large division problems resulted in the creation of three distinct strategy groups: *Retrievers* (*n* = 13), *Transformers* (*n* = 15), and *counters* (*n* = 6). A breakdown of procedure selection for each group is presented in Table 2.

Table 2

Experiment 1A: Summary of procedures reported (in%) for retrievers, transformers, and counters.

	Groups					
	Retrievers		Transformers		Counters	
Procedures	*M*	*SD*	*M*	*SD*	*M*	*SD*
Memory	77	14	39	19	33	14
Recast	18	16	58	21	20	18
Count	4	9	3	5	47	20

Traditional Analyses

A 2 (Problem Size [small, large]) x 3 (Group [retrievers, transformers, counters]) mixed-factors ANOVA was performed on data for each traditional dependent measure, with group as the between-subjects factor and problem size as the within-subjects factor. Results of these ANOVAs are displayed in Table 3.

Mean Response Time. Participants were slower on large problems than on small problems (2105 vs. 1424 ms). Counters were the slowest (2451 ms), followed by transformers (1554 ms), then retrievers (1288 ms). Post hoc tests revealed no significant difference in RT between retrievers and transformers ($p = .30$). In contrast, both retrievers and transformers showed smaller RTs than counters ($p < .001$). There was also a problem size x group interaction. Counters (1166 ms, *CI* = 237 ms) and transformers (574 ms, *CI* = 151 ms) showed a significant problem-size effect, but retrievers (303 ms, *CI* = 161 ms) did not.

Standard Deviation. Participants' latencies were more variable on large problems than on small problems (1094 vs. 677 ms). Counters' latencies varied the most (1424 ms), followed by transformers (726 ms) and retrievers (507 ms). Post hoc tests revealed no significant variability in latencies between retrievers and transformers ($p = .30$), but counters were more variable than both retrievers and transformers ($p < .001$). Unlike mean response time, there was no group x problem size interaction, suggesting that each group was more variable on large than on small problems.

Table 3

Experiment 1A: Summary of separate 2 (problem size: small, large) by 3 (group: retrievers, transformers, counters) ANOVAs for traditional measures of performance.

		Dependent Variables					
		Reaction Time		Standard Deviation		Accuracy	
Independent Variables	*df*	*F*	η_p^2	*F*	η_p^2	F	η_p^2
Problem Size	1,31	82.68***	.73	51.28***	.62	37.09***	.55
Group	1,31	16.34***	.51	15.16***	.49	1.13	.07
Problem Size x Group	2,31	9.40**	.38	.83	.05	4.01*	.21

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Accuracy. Participants were less accurate on large problems than small problems (87 vs. 95%). There was no main effect of group, such that performance was similar for retrievers (93%), transformers (91%), and counters (88%). The problem size x group interaction was significant. Counters showed the largest problem-size effect (14%), followed by transformers (8%) and retrievers (4%).

Ex-Gaussian Analyses

A 2 (Problem Size [small, large]) x 3 (Group [retrievers, transformers, counters]) mixed-factors ANOVA was performed on both mu and tau values, with group as the between-subjects factor and problem size as the within-subjects factor. Results of these ANOVAs are displayed in Table 4.

Model Fits. In order to determine the appropriateness of the ex-Gaussian model, correct trials for each participant were separated into two distinct vectors of data (small problems and large problems). These vectors were then fitted using specialized Matlab source code that calculates the value for the Akaike Information Criterion (AIC; Helie, 2006). Specifically, the AIC estimates the covariance between the observed data and the model's predictions using (1) the minus log-likelihood value and (2) the number of parameters. Simply put, AIC values are compared to determine which model provides the best overall fit for the observed data set. Smaller AIC values indicate overall better fit.

Thus, although the AIC does not specifically evaluate whether the model is a good fit for the data, it does provide a means for comparing and selecting the models that minimize the loss of data. The ex-Gaussian distribution was compared to three competing models: the normal, exponential, and lognormal distributions. Of particular interest was the comparison between the lognormal and ex-Gaussian fits, as the lognormal distribution has also been used to characterize RT distributions (Ratcliff & Murdock, 1976). Δ AIC values were calculated by subtracting the minimum AIC (in most cases, the ex-Gaussian fits) from the AIC values for each competing model. The ex-Gaussian model provided a better fit of the data for 65 of the total 68 fits when compared to the normal distribution. Δ AIC values revealed poor support overall for the accuracy of the normal distribution. For all 68 fits, the ex-Gaussian model was a better fit than the exponential model. Again, Δ AIC values suggested extremely poor support of the exponential distribution to accurately fit RT distributions. Finally, the ex-Gaussian was a better overall fit of RT distributions than the lognormal for 60 of 68 fits. Fits for which the ex-Gaussian was not ideal were still included in the data set, as these values still provide important information concerning the shape of the distributions (Campbell & Penner-Wilger, 2006; Penner-Wilger et al., 2002).

Table 4

Experiment 1A: Summary of separate 2 (problem size: small, large) by 3 (group: retrievers, transformers, counters) ANOVAs for ex-Gaussian parameters.

		Dependent Variables			
		Mu		Tau	
Independent Variables	*df*	*F*	η_p^2	*F*	η_p^2
Problem Size	1,31	13.86***	.31	92.68***	.75
Group	2,31	3.38**	.18	17.88***	.54
PS x Group	2,31	3.20*	.17	8.15***	.34

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Mu. The mean of the leading edge component was greater for large problems than for small problems (1014 vs. 870 ms). The main effect of group was also significant. Mu was smallest for retrievers (863 ms), followed by transformers (902 ms) and counters

(1061 ms). Post hoc tests revealed that retrievers had smaller mu values than counters ($p < .05$), but no other comparisons were significant. Finally, the interaction of problem size and group approached significance. As shown in Figure 3, counters showed a significant problem-size effect (282 ms, CI = 122 ms), but transformers (126 ms, CI = 78 ms) and retrievers (24 ms, CI = 84 ms) did not. Accordingly, mu values did not vary across groups on small problems, whereas on large problems, counters had large mu values than transformers or retrievers.

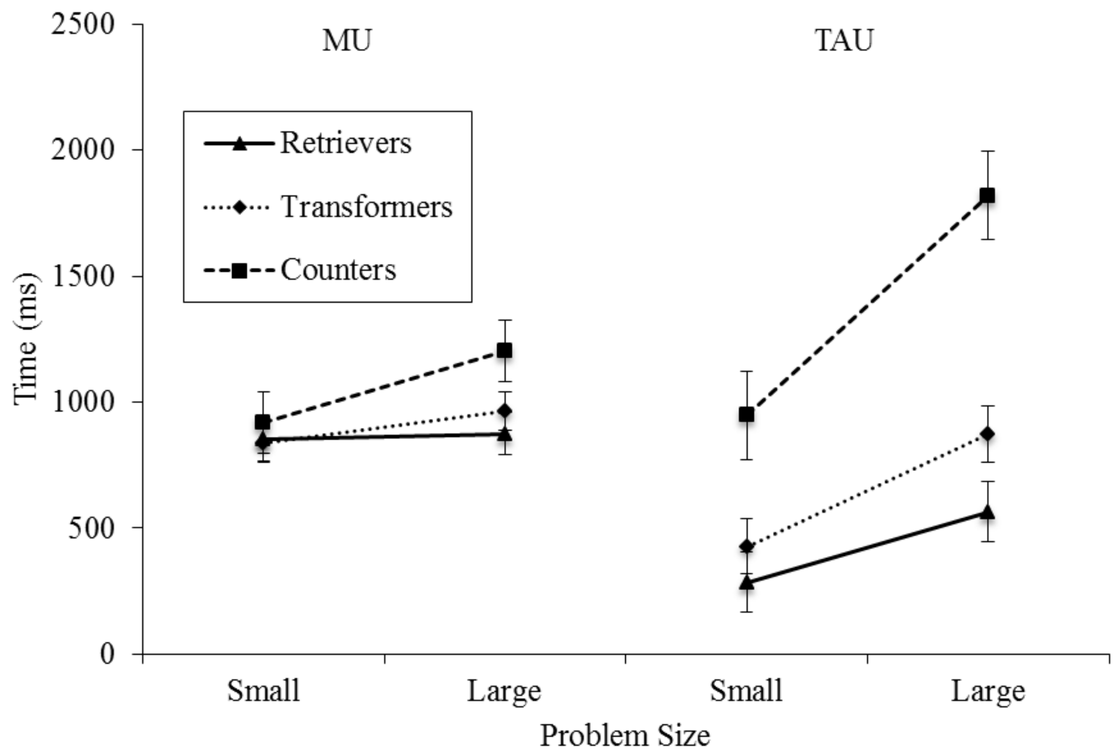

Figure 3. Experiment 1A: Average mu and tau values (ms) for retrievers, transformers, and counters as a function of problem size. Group membership based on self-reports.

Tau. The mean of the exponential component was greater for large problems than small problems (1086 vs. 554 ms). Counters had the largest tau values (1383 ms), followed by transformers (651 ms) and retrievers (425 ms). Post hoc tests revealed that both retrievers and transformers had smaller tau values than counters ($p < .001$), but retrievers and transformers did not differ from one another ($p = .23$). Finally, problem size interacted with group. As shown in Figure 3, counters showed the largest problem-size effect (871 ms, *CI* = 176 ms), followed by transformers (446 ms, *CI* = 110 ms) and retrievers (279 ms, *CI* = 118 ms). Retrievers and transformers were not different for tau on small problems, whereas counters had large tau values. On large problems, tau values varied across all three groups.

Fixation Analyses

A 2 (Problem Size [small, large]) x 3 (Group [retrievers, transformers, counters]) x 3 (Interest Area [left operand, center, right operand]) mixed-factors ANOVA was performed on both gaze duration and fixation count values, with group as the between-subjects factor and problem size and interest area as within-subjects factors. Results of these ANOVAs are displayed in Table 5.

Table 5

Experiment 1A: Summary of separate 2 (problem size: small, large) by 3 (interest area: left operand, operator, right operand) by 3 (group: retrievers, transformers, counters) ANOVAs for gaze duration and number of fixations.

		Dependent Variables			
		Gaze Duration		Fixation Count	
Independent Variables	*df*	*F*	η_p^2	*F*	η_p^2
Problem Size	1,31	70.89***	.70	79.22***	.72
Interest Area	2,62	8.51**	.22	7.77**	.20
Group	2,31	17.76***	.53	10.27***	.40
PS x Group	2,31	8.95**	.37	5.15*	.25
IA x Group	4,62	4.08**	.21	2.53*	.14
PS x IA	2,62	62.59***	.67	74.66***	.71
PS x IA x Group	4,62	5.35**	.26	3.93**	.20

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Gaze Duration. Participants looked longer at large problems than small problems, (495 vs. 321 ms). Consistent with the overall latency analysis, counters had the longest gaze durations (595 ms), followed by transformers (348 ms) and retrievers (281 ms). Counters' gaze durations were significantly longer than those of retrievers and transformers ($p < .001$), but retrievers and transformers did not differ from each other ($p = .33$). Participants looked at each interest area for unequal periods (left operand = 547 ms; operator = 335 ms; right operand = 341 ms). In particular, participants looked at the left operand for longer than at the operator or right operand ($p < .01$), but looked at the operator and right operand for equally ($p = 1.00$). The problem size x interest area interaction was also robust. As problems increased in size, participants looked less at the operator and more at the operands. Finally, there was a significant group x problem size x interest area interaction, indicating that the problem size x interest area interaction varied across groups. As shown in Figure 4, counters and transformers began looking much more at the operands as a function of problem size when compared to retrievers. This pattern was especially evident on the left operand, but was also found for the right operand. Time spent on the operator did not differ across problem size for any group.

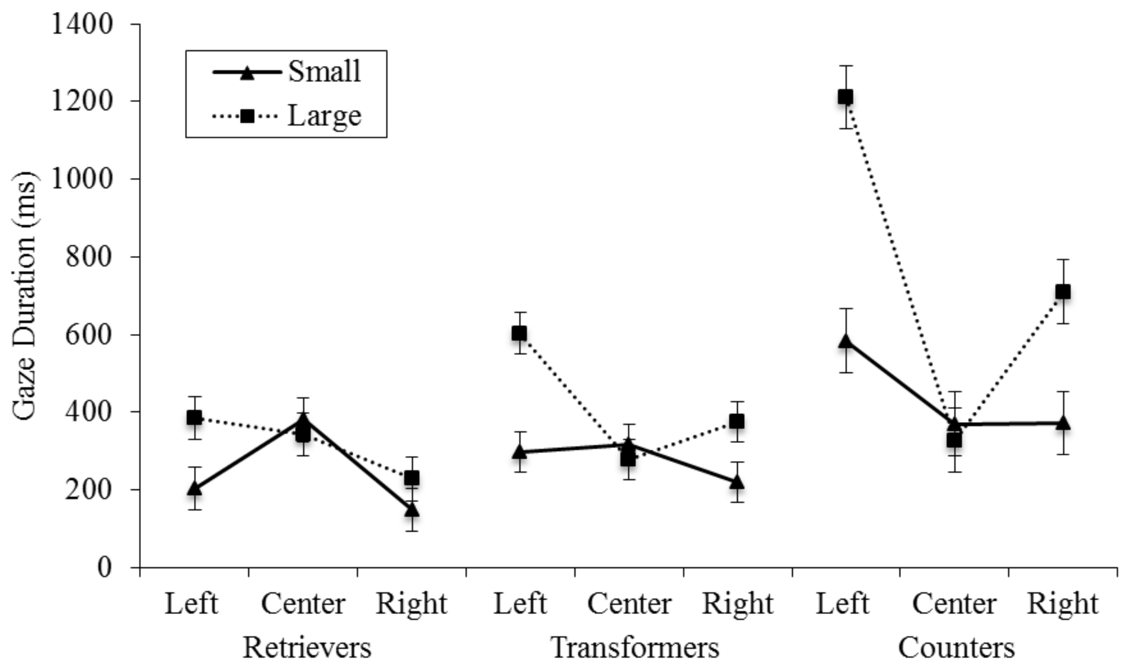

Figure 4. Experiment 1A: Mean gaze duration (ms) for small and large problems as a function of procedure group and interest area. Group membership based on self-reports.

Fixation Count. Participants made more fixations on large problems than on small problems (1.73 vs. 1.27), $F(1, 31) = 79.22$, $p < .001$, $\eta p^2 = .72$. The number of fixations made was not evenly distributed across problem components (left operand = 1.87; operator = 1.44; right operand = 1.18). Participants fixated more on the left operand than the right operand ($p < .001$) or the operator ($p = .07$), but made equal fixations on the operator and right operand ($p = .56$). Counters made the largest number of fixations (1.96), followed by transformers (1.42) and retrievers (1.11). Counters made more fixations than retrievers ($p < .001$) and transformers ($p < .05$), but retrievers and transformers did not differ from each other ($p = .12$). There was also a significant problem size x interest area interaction. Participants made the majority of their fixations on the operator area for small problems, but started making more fixations on the operands for large problems. Finally, the group x problem size x interest area interaction was significant. As shown in Figure 5, counters and transformers made more fixations on the operands as a function of problem size when compared to retrievers. The number of fixations on small and large problems did not vary and was similar across groups for the operator.

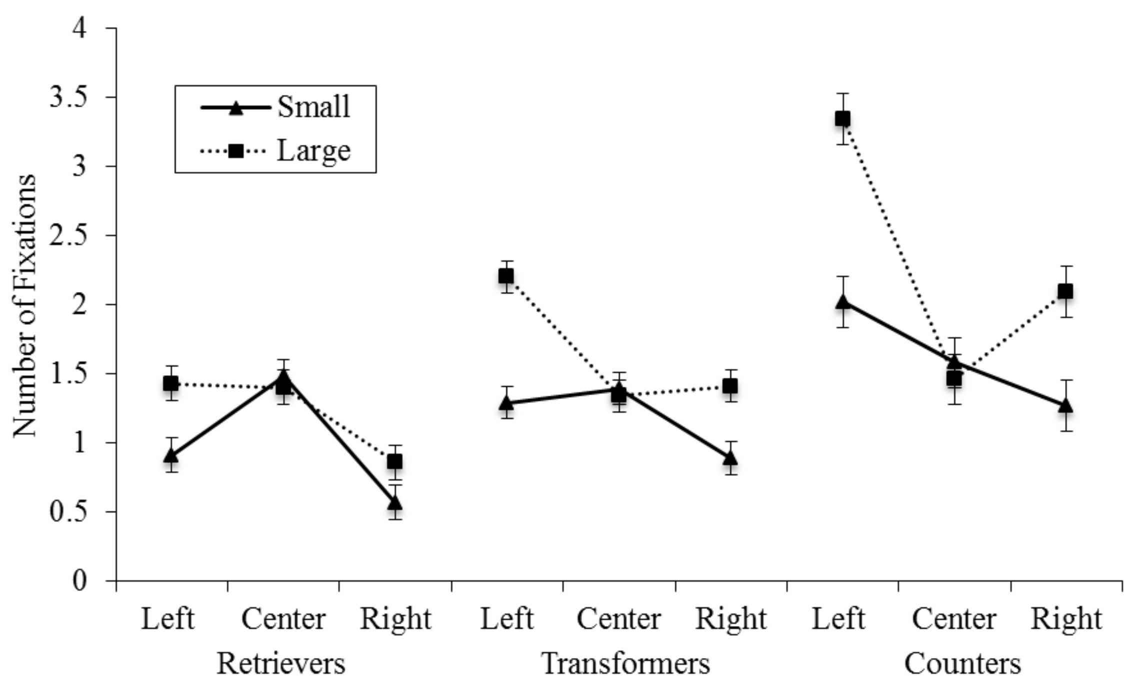

Figure 5. Experiment 1A: Mean number of fixations for small and large problems as a function of procedure group and interest area. Group membership based on self-reports.

Summary

The first goal of Experiment 1A was to replicate previous research that used eye movements to assess division performance but where self-reports were not collected (Curtis et al., 2014). As expected, patterns of gaze and fixation counts mirrored those reported in previous findings. Participants fixated primarily at the operator when solving small problems, but began fixating more on the left and right operands for large problems. The second goal of Experiment 1A was to isolate and explore individual differences regarding procedural solutions. Three groups (retrievers, transformers, and counters) were created based on participants' self-reports for large division problems.

Retrievers used memory to solve the majority of large problems, transformers primarily relied on recasting, and counters showed the greatest variability in their selection of procedures, with an emphasis on repeated subtraction to solve the problems. Complementing the results from previous research (LeFevre et al., 2006; Penner-Wilger et al., 2002), patterns of mu and tau appeared to be systematically related to participants' descriptions of solution procedures. In particular, counters and transformers showed a larger problem-size effect in tau than retrievers. Counting and recasting are arguably slower and less accurate than directly retrieving an answer from memory, as they require additional mental steps to calculate the correct solution. These procedures make up the slower responses located within the skew of the RT distribution. Thus, increased use of slow, non-retrieval procedures likely corresponds to increases in tau (Penner-Wilger et al., 2002).

In contrast to tau, counters were the only group who showed a significant problem-size effect in mu. Campbell and Penner-Wilger (2006) proposed that increases in mu plausibly represent a global shift from using retrieval on small problems to non-retrieval procedures on large problems. Counters reported the greatest overall use of procedures (67%) compared to transformers (61%) and retrievers (22%) on large problems. Thus, increases in mu might reflect the overall extent to which participants rely on procedures vs. retrieval.

Fixation analyses revealed substantially different patterns of performance among retrievers, transformers, and counters. In particular, retrievers looked primarily at the operator for small problems and only started looking more at the left operand for large problems. Transformers looked primarily at the center for small problems but started looking more at the operands for large problems, with a heavier emphasis placed on the left operand. Finally, counters primarily looked at the left operand for both small and large problems. These results appear to show that when solvers implement different solution procedures they require different information from various problem components.

Experiment 1B

Although it is tempting to conclude that eye movements reflect different solution procedures, the results of Experiment 1A were based on participants' self-reports. As mentioned previously, some researchers have claimed that self-reports influence overall patterns of performance (Fayol & Thevenot, 2012; Kirk & Ashcraft, 2001). As such, the observed patterns above may not accurately reflect performance when self-reports are absent from the experimental design. Thus, the goal of Experiment 1B was to find alternate group classification criteria. Researchers have claimed that ex-Gaussian parameter estimates provide a more objective way of assessing individual differences regarding simple arithmetic performance (LeFevre et al. 2006; Penner-Wilger et al.

2002), especially when self-reports are not used. Considering that ex-Gaussian parameter estimates logically complement participants' descriptions of solution procedures, participants who are classified based on mu and tau values should elicit similar eye movement patterns as observed in Experiment 1A.

The 32 participants who did not report their solution procedures completed a total of 6528 trials. Evaluation of the raw data revealed 413 (6%) trials that were spoiled due to equipment failure (i.e., failing to detect an utterance or firing prematurely) and a further 658 (10%) trials in which participants responded incorrectly. The remaining 5457 valid trials, along with the 5,530 valid trials from Experiment 1A, were then considered for group categorization based on participants' mu and tau values for large problems. Bonferroni post hoc tests were used to compare main effects and 95% confidence intervals (*CI*s) were used for the interpretation of interactions (Jarmasz & Hollands, 2009).

Discriminant Function Analysis

Participants' mu and tau values for large division problems were used in a discriminant function analysis as independent variables to categorize retrievers and transformers. Participants in both the self-report and no self-report conditions were included in the analysis to determine how well the reported use of retrieval on large division problems correlated to ex-Gaussian parameter estimates. The discriminant function analysis correctly categorized 68% of the original grouped cases: 10 of 13 retrievers, 7 of 15 transformers, and 6 of 6 counters. For the participants who did not provide self-reports, the analysis identified 15 as retrievers, 12 as transformers, and 7 as counters. This resulted in 31 retrievers, 22 transformers, and 15 counters in total.

Traditional Analyses

All of the 68 participants were included in all subsequent analyses. A 2 (Problem Size [small, large]) x 3 (Group [retrievers, transformers, counters]) mixed-factors ANOVA was performed on data for each traditional dependent measure, with group as the between-subjects factor and problem size as the within-subjects factor. Results of these ANOVAs are displayed in Table 6.

Mean Response Time. Participants were slower on large problems than on small problems (2032 vs. 1395 ms). The main effect of group was also significant. Counters made the slowest responses (2360 ms), followed by transformers (1600 ms) and retrievers (1179 ms). Post hoc tests revealed that all groups were significantly different from one another ($p < .001$). There was also a problem size x group interaction. Counters (1131 ms, $CI = 80$ ms) showed the largest problem-size effect, followed by transformers (537 ms, $CI = 66$ ms) and retrievers (243 ms, $CI = 56$ ms).

Standard Deviation. Participants' latencies were more variable on large problems than on small problems (1067 vs. 618 ms). There was also a main effect of group. Counters showed the greatest amount of variability (1376 ms), followed by transformers (753 ms) and retrievers (397 ms). Post hoc tests revealed that all groups were significantly different from one another ($p < .001$). There was also a problem size x group interaction. Counters showed the largest problem-size effect (643 ms, $CI = 90$ ms), followed by transformers (510 ms, $CI = 74$ ms) and retrievers (194 ms, $CI = 62$ ms).

Accuracy. Participants were less accurate on large problems than on small problems (84 vs. 95%). The main effect of group was also significant. Counters were the least accurate (84%), followed by transformers (92%) and retrievers (93%). Post hoc tests revealed that retrievers and transformers performed the same ($p = 1.00$). In contrast, counters were less accurate than both retrievers and transformers ($p < .001$). There was also a problem size x group interaction. Counters showed the largest problem-size effect (16%, *CI* = 3%), followed by transformers (9%, *CI* = 2.5%) and retrievers (6%, *CI* = 2.1%).

Table 6

Experiment 1B: Summary of separate 2 (problem size: small, large) by 3 (group: retrievers, transformers, counters) ANOVAs for traditional measures of performance.

		Dependent Variables					
		Reaction Time		Standard Deviation		Accuracy	
Independent Variables	*df*	*F*	η_p^2	*F*	η_p^2	F	η_p^2
Problem Size	1,65	514.62***	.89	207.12***	.76	96.18***	.60
Group	2,65	121.40***	.79	134.24***	.81	11.34***	.26
Problem Size x Group	2,65	81.15***	.71	20.30***	.38	8.03**	.20

Note. ** $p < .01$. *** $p < .001$.

Ex-Gaussian Analyses

Separate 2 (problem size [small, large]) x 3 (group [retrievers, transformers, counters]) mixed-factors ANOVAs were performed on mu and tau values, with group as the between-groups factor and problem size as the within-groups factor. Results of these ANOVAs are displayed in Table 7.

Table 7

Experiment 1B: Summary of separate 2 (problem size: small, large) by 3 (group: retrievers, transformers, counters) ANOVAs for ex-Gaussian parameters.

		Dependent Variables			
		Mu		Tau	
Independent Variables	*df*	*F*	η_p^2	*F*	η_p^2
Problem Size	1,65	47.86	.42	543.05	.89
Group	2,65	15.62	.33	181.06	.85
PS x Group	2,65	11.82	.27	75.05	.70

Note. All obtained F ratios are significant at the .001 level.

Model Fits. Appropriateness of the ex-Gaussian model for RT distributions was determined using the same methodology for model fits in the self-report condition. In this case, model fits were conducted for the 34 participants who did not provide self-reports. Comparisons between the ex-Gaussian and normal distributions revealed that the ex-Gaussian was a better fit of the data for all 64 of the 68 total fits. Evaluation of Δ AIC values revealed that the normal distribution was highly unlikely to fit the data best. Likewise, the ex-Gaussian was a better fit for all 68 RT distributions when compared to the exponential model. Again, Δ AIC values revealed that the exponential model supported the data extremely poorly. Finally, the ex-Gaussian provided the best fit for 63 of the 68 RT distributions when compared to the lognormal model. Distributions for which the ex-Gaussian was not the best fitting model were still included in the analysis.

Mu. The mean of the leading edge component was greater for large problems than for small problems (1007 vs. 875 ms). The main effect of group was significant. Counters had the largest mu values (1042 ms), followed by transformers (945 ms) and retrievers (836 ms). Post hoc tests revealed that retrievers had significantly smaller mu values than counters and transformers ($p < .001$). The difference between transformers and counters approached significance ($p = .06$). Finally, the interaction of problem size and group was significant. As displayed in Figure 6, counters showed a significant problem-size effect (274 ms, CI = 55 ms), but transformers (72 ms, CI = 46 ms) and retrievers (51 ms, CI = 38 ms) did not.

Tau. The mean of the exponential component was greater for large problems than for small problems (1041 vs. 520 ms). The main effect of group was also significant. Counters showed the largest tau values (1346 ms), followed by transformers (650 ms) and retrievers (345 ms). Post hoc tests revealed that all group differences were significant ($p < .001$). Finally, problem size interacted with group. As displayed in Figure 6, counters showed the largest problem-size effect (864 ms, CI = 64 ms), followed by transformers (502 ms, CI = 53 ms) and retrievers (194 ms, CI = 45 ms).

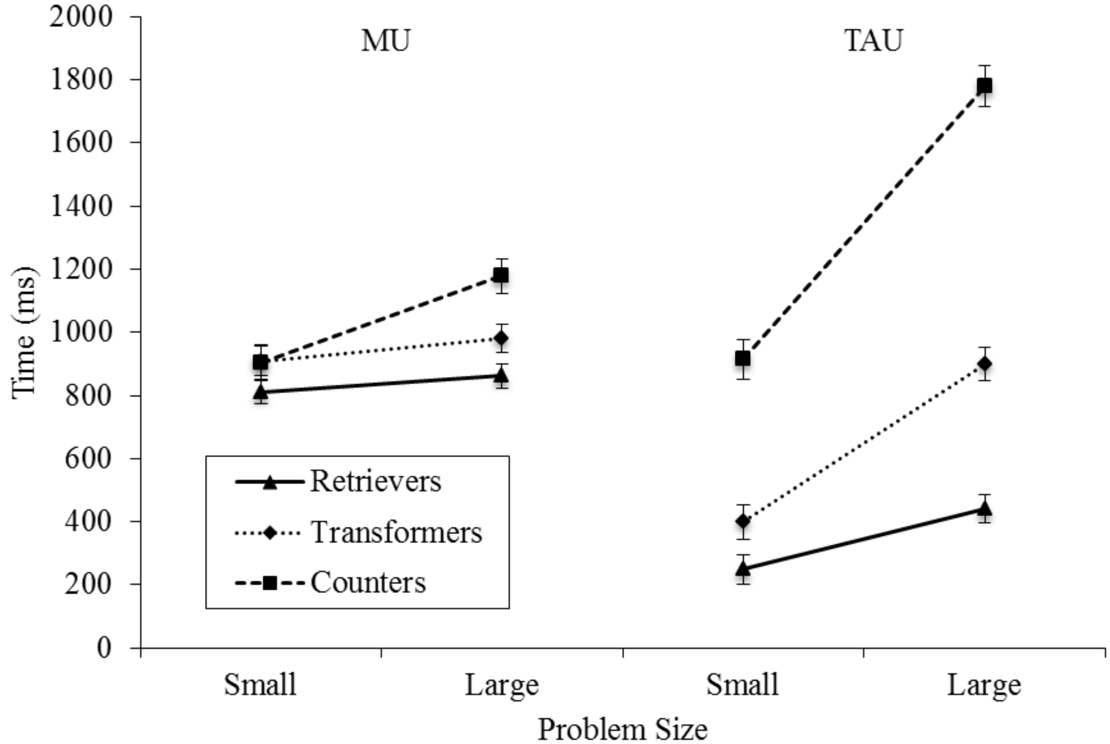

Figure 6. Experiment 1B: Average mu and tau values (ms) for retrievers, transformers, and counters as a function of problem size. Group membership is based on mu and tau values for large division problems.

Fixation Analyses

A 2 (problem size [small, large]) x 3 (group [retrievers, transformers, counters]) x 3 (interest area [left operand, operator, right operand]) mixed-factors ANOVA was performed on both gaze duration and fixation count values, with group as the between-subjects factor and problem size and interest area as within-subjects factors. Results of these ANOVAs are displayed in Table 8.

Table 8

Experiment 1B: Summary of separate 2 (problem size: small, large) by 3 (interest area: left operand, operator, right operand) by 3 (group: retrievers, transformers, counters) ANOVAs for gaze duration and number of fixations.

		Dependent Variables			
		Gaze Duration		Fixation Count	
Independent Variables	*df*	*F*	η_p^2	*F*	η_p^2
Problem Size	1,65	362.85	.85	504.59	.89
Interest Area	2,130	19.49	.23	24.69	.28
Group	2,65	54.78	.63	62.00	.66
PS x Group	2,65	58.94	.65	72.43	.69
IA x Group	4,130	8.43	.21	8.09	.20
PS x IA	2,130	151.33	.70	164.27	.72
PS x IA x Group	4,130	20.33	.39	18.51	.36

Note. All obtained F ratios are significant at the .001 level.

Gaze Duration. Participants looked at large division problems longer than at small division problems (481 vs. 321 ms). Gaze duration was not divided evenly across problem components (left operand = 540 ms; operator = 358 ms; right operand = 305 ms). Participants looked at the left operand more than the operator or right operand ($p < .001$), but looked at the operator and right operand for equal periods of time ($p = .57$). The main effect of group was also significant. Counters looked at problems longest (563 ms), followed by transformers (370 ms) and retrievers (270 ms). Gaze durations for all three groups were significantly different ($p < .001$). The problem size x interest area interaction was also robust. As problems increased in size, participants looked less at the center and more at the left and right operands. Finally, the group x problem size x interest area interaction achieved statistical significance. As shown in Figure 7, the problem size x interest area interaction varied across groups. Specifically, counters and transformers looked much more at the operands as problem size increased when compared to retrievers.

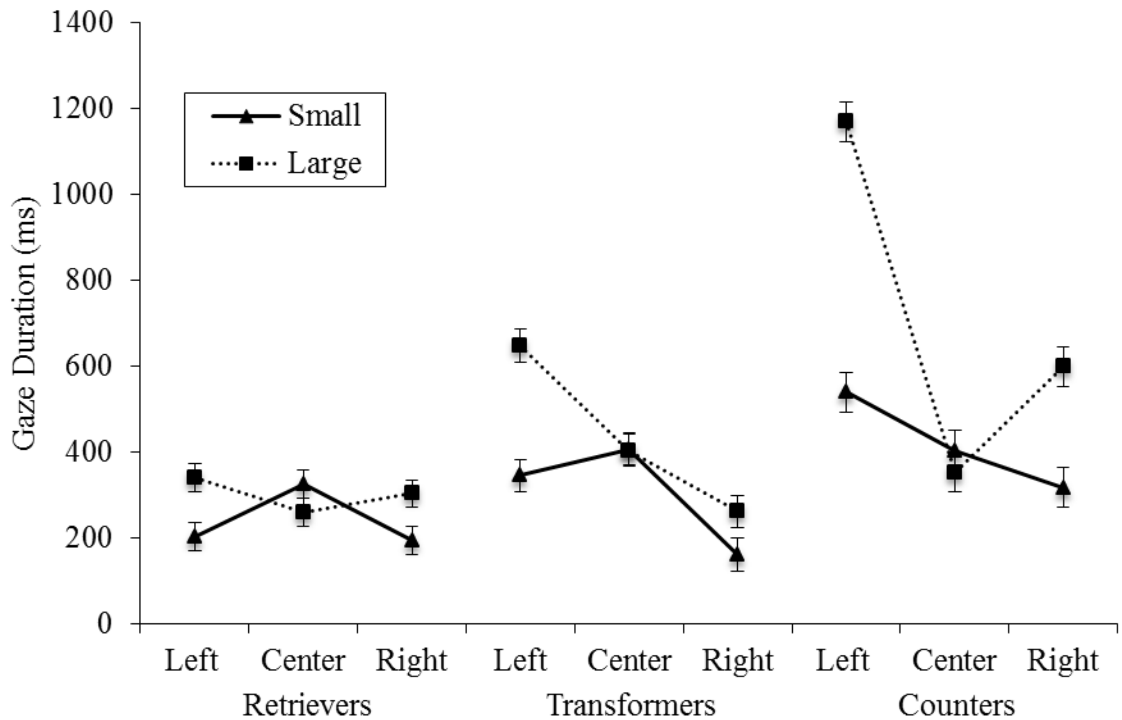

Figure 7. Experiment 1B: Mean gaze duration (ms) for small and large problems as a function of procedure group and interest area. Group membership is based on mu and tau values for large division problems.

Fixation Count. Participants made more fixations on large problems than on small problems (1.78 vs. 1.30). The number of fixations made was not evenly distributed across problem components (left operand = 1.96; operator = 1.49; right operand = 1.16). All three interest areas differed from one another ($p < .05$). Counters made the most fixations (2.10), followed by transformers (1.41) and retrievers (1.11), with all three groups differing significantly from one another ($p < .001$). There was also a significant problem size x interest area interaction. Participants made the majority of their fixations on the operator for small problems, but made more fixations on the operands for large problems. Finally, the group x problem size x interest area interaction was significant. As shown in Figure 8, the problem size x interest area interaction varied across groups. Specifically, counters and transformers made more fixations on the operands as problem size increased when compared to retrievers.

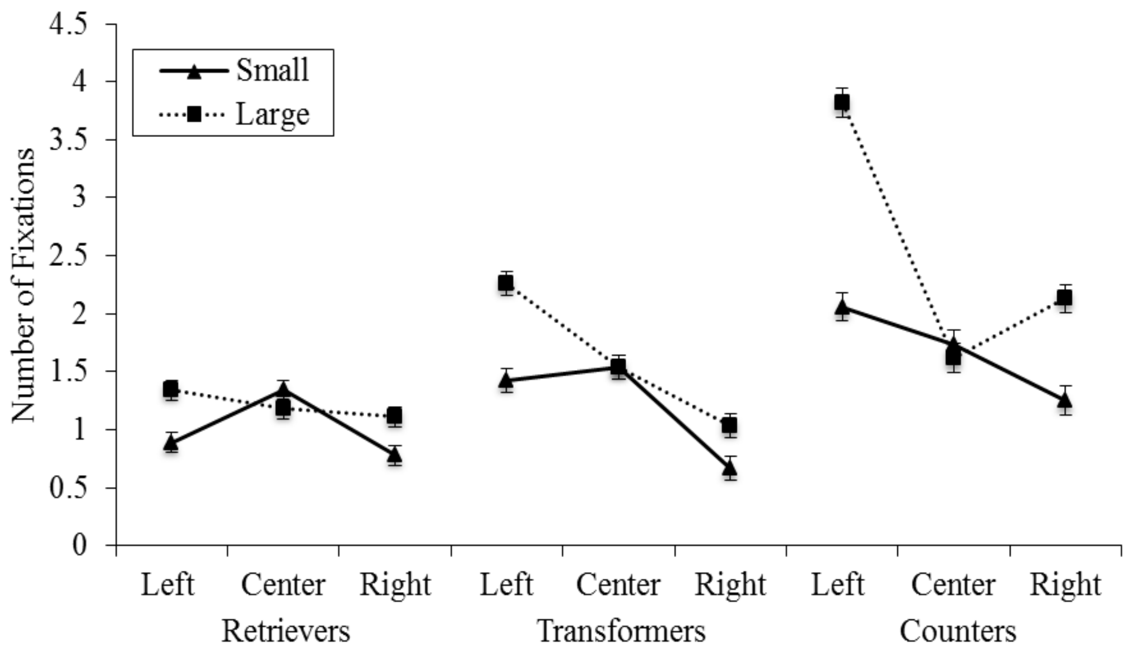

Figure 8. Experiment 1B: Mean number of fixations for small and large problems as a function of procedure group and interest area. Group membership is based on mu and tau values for large division problems.

Summary

The primary goal of Experiment 1B was to assess whether the observed differences in eye-movement behavior across procedure groups could be detected in the absence of self-reports. Historically, researchers have advocated the use of ex-Gaussian parameter estimates for RT distributions in order to classify participants into different procedure groups (LeFevre et al., 2006; Penner-Wilger et al., 2002). Thus, mu and tau values for large division problems were used in a discriminant function analysis to examine the extent to which properties of RT distributions complement more traditional measures of performance.

The discriminant function analysis was fairly accurate in correctly classifying participants who provided self-reports. Ex-Gaussian analyses revealed similar patterns of performance for mu and tau as witnessed in Experiment 1A. Furthermore, the three procedure groups, as classified by using mu and tau values for large problems, elicited similar patterns of performance in the form of gaze duration and fixation counts as the groups determined solely by self-reports. Thus, the self-report and the ex-Gaussian analyses were complementary.

Discussion

Understanding the mental processes that underlie basic arithmetic performance is an important goal for researchers of mathematical cognition. Based on a wide variety of empirical studies that utilized a self-report methodology, adults appear to solve even the most basic arithmetic problems in a variety of ways. One explanation of the problem-size effect is that participants use memory retrieval on small arithmetic problems whereas they use a variety of procedures on large problems. Although self-reports appear to complement other measures of performance such as mean RT and variability, their use has not gone unchallenged. For instance, performance appears to vary depending on instructional demand and arithmetic skill levels (Smith-Chant & LeFevre, 2002). As a result, advocates of traditional, introspective methods have been forced to provide converging evidence to support the use of self-reports in understanding arithmetic performance. One alternative is to examine RT distributions and assess changes in parameter values as a function of problem size and skill group (Campbell & Penner-Wilger, 2006; LeFevre et al., 2006). Another, more novel approach, is to examine changes in eye-movement behavior as a function of problem size and differences in procedure selection (Curtis et al., 2014; Huebner et al., 2013). The goal of the present research was to utilize both alternatives to augment traditional, introspective methods of assessing cognitive processing when adults solved basic division facts.

All predictions concerning patterns of performance were supported in the first half of the experiment. Overall, participants showed a typical problem-size effect in terms of mean response time and error rate. Analyses of overall gaze duration and fixation counts revealed performance patterns identical to those reported for division by Curtis et al. (2014), indicating that the inclusion of self-reports did not influence patterns of behavior.

An examination of self-reports revealed that participants varied in their selection of procedures, especially for large division problems. As such, participants were classified into three procedure groups in a similar manner to previous studies that examined individual differences with subtraction. Similar to the results reported by LeFevre et al. (2006), all three groups showed differential patterns of mu and tau. Counters showed the largest problem-size effect in tau compared to the other two procedure groups. Counting back by the divisor to reach the correct solution naturally requires more steps than simply relying a derived multiplication fact, and each step may also require more time. As such, tau might be a useful index for identifying individuals who count on large problems, irrespective of operation. Retrievers showed a negligible shift in mu across problem size and only a small increase in tau when compared to transformers and counters. The lack of a shift in mu possibly reflects consistent use of direct retrieval to solve basic division facts across problem size. The slight increase in tau may have reflected the tendency for retrievers to use some transformations on large division problems, as evidenced by their self-reports. Transformers also showed no problem-size effect in mu, but showed a larger increase in tau than retrievers. Again, this larger increase in tau plausibly represents greater reliance on time-consuming procedures as problem size increases.

These patterns complement the findings of Robinson et al. (2002), who noted substantial differences in response latencies on division problems between individuals who primarily used retrieval and individuals who primarily used transformations and counting. These findings also complement the results reported by LeFevre et al. (2006), who showed no differences with problem size for mu for their highly skilled group of participants on subtraction problems, but showed major differences across problem size for both mu and tau for their low-skilled participants.

Retrievers paid more attention to the operator for small problems and began looking more left operand for large problems. In contrast, transformers switched from looking at the operator for small problems to looking at the left and right operands for large problems, with a heavier emphasis placed on the left operand. Counters showed the greatest deviations in performance patterns when compared to retrievers and transformers, looking at the left operand more frequently than at the operator for small problems and looking much more at the left and right operands than retriever and transformers for large problems. The heavier emphasis placed on the operands by counters for small problems may reflect their tendency to use a higher rate of procedures on small problems compared to retrievers and transformers. These findings support the reasoning that different procedures require different information contained within specific problem components.

Based on previous findings that investigated performance across multiplication and division (LeFevre & Morris, 1999; Mauro et al. 2003), the heavier emphasis placed on the operands by transformers may represent a mental rearrangement of the problem to utilize multiplication knowledge (e.g., rearranging 56 / 8 to 8 x ? = 56). The similar, yet larger number of fixations placed on the operands by counters may represent their tendency to count back by the divisor and periodically check the value of the dividend. In contrast, direct memory retrieval arguably does not require manipulations or extensions of the numbers to reach a correct solution, explaining the tendency for retrievers to rely less on the operands, even for larger problems.

The discriminant function analysis was fairly accurate in classifying the participants who provided self-reports and the resulting analyses showed similar patterns across groups. Retrievers showed a negligible shift in mu and only a small increase in tau. Transformers also showed a negligible shift in mu but a greater increase in tau than retrievers, again reflecting the assertion that more frequent use of time-consuming procedures builds in the tail of the RT distribution. This finding provides further empirical support for the use of ex-Gaussian parameter estimates as complementary measures to participants' descriptions of their solutions.

In conjunction with the ex-Gaussian analysis, eye-movement patterns across skill groups were fairly similar to those found in the self-report condition. One notable difference comes from patterns of gaze duration in which retrievers began focusing more on the left *and* right operands for large problems relative to the operator. As noted by LeFevre et al. (2006), the discriminant function analysis resulted in similar--not identical--results when mu and tau values were used to classify participants instead of self-reports.

Why, then, might the discriminant function not provide more accurate categorization, particularly for the group that self-reports indicated were transformers? One potential explanation comes from the work of Imbo, Vandierendonck, and Rosseel (2007) and Imbo and Vandierendonck (2008), who specifically discriminate between procedure selection and procedure efficiency. In their research, participants solved basic addition and multiplication problems across varying numbers of practice sessions. The authors found that participants relied more frequently on retrieval with increased practice. Interestingly, however, both retrieval trials and non-retrieval trials showed an increase in overall efficiency as a result of increased practice sessions. The authors concluded that procedures such as transformation and counting become more efficient with increased practice and that highly skilled adults utilize these procedures more efficiently in general than less-skilled participants. Thus, if the transformers in Experiment 1A used procedures such as recasting with relatively high efficiency, it may explain why the discriminant function analysis classified 6 of them as retrievers. Moreover, the increased attention paid to the operands by retrievers may still reflect the occasional use of a procedure other than retrieval, but these procedures are executed with greater efficiency than participants who more readily use recasting and counting with less efficiency.

Finally, how might the experimental design used here translate to another operation of interest? Huebner and LeFevre (2014) conducted a second subtraction experiment in which participants provided immediate, retroactive descriptions of their solution procedures. This allowed us to further investigate changes in eye patterns both as a function of problem size and the selection of procedures across groups. Three performance groups were created based on participants' self-reports on large problems: retrievers, transformers, and counters. Again, if direct retrieval does not require any mental rearrangement or manipulation of the operands to solve the problem, we predicted that individuals who rely primarily on memory retrieval for large subtractions would allocate their attention primarily to the operator. In contrast, transformers and counters should allocate more attention to the operands as a function of problem size. As predicted, retrievers primarily looked at the center of the problem for both small and large subtractions. In contrast, transformers and counters switched from focusing primarily on the center for small problems to focusing primarily on the operands for large problems. Notably however, in subtraction counters spent more time on the right operand whereas in division, more time was spent on the left operand. Why might counters' allocation of attention differ between operations? While both division and subtraction problems can be solved by counting back, subtraction is usually solved by counting back in single-unit increments. For instance, given the problem 12 – 3, counters may extend the right operand into units of one and count back until 9 is reached. In this case, the right operand receives more attention. In contrast, a division problem such as 24 ÷ 4 may require a counter to extend the left operand into 6 equal parts in order to reach 6, therefore allocating more attention to the left operand. In any case, available procedures do not appear to translate perfectly across operations.

Limitations

While the results of the present study show potential for eye tracking as a useful tool by which to measure arithmetic performance, a number of limitations need to be addressed. First, eye tracking is rarely used in the domain of mathematical cognition and, as a result, standardized measures of fixations are lacking in the available literature. For example, Green et al. (2007) admitted that their choice of minimum fixation time (60 ms) was arbitrary. Furthermore, Verschaffel et al. (1994) used a minimum time of 100 ms to mark when a fixation was produced. The present study classified a fixation whenever participants' eyes fell within a particular interest area for 50 ms. Therefore, it is completely possible that altering my definition of a fixation could result in different performance patterns. Standardized methodologies should be promoted if the use of eye movements to assess procedure use in arithmetic studies increases.

Second, it remains relatively unclear how best to describe procedure use to participants. For example, Campbell and Timm (2000) examined retrieval interference effects by having participants either solve division and addition problems in one experimental condition or multiplication and addition problems in a second condition. The authors used the same methodology here in which participants were provided specific descriptions and examples of common procedures. This method is different from that used by LeFevre and Morris (1999), in which participants were not given specific examples. Campbell and Timm (2000) found that their participants reported memory retrieval on 90% of division problems, whereas the participants from the LeFevre and Morris (1999) study reported memory retrieval on 45% of division problems. While the discrepancies between the two studies may be attributed solely to population or sampling differences, variations in instructional demand can never be discounted (Kirk & Ashcraft, 2001). Thus, providing participants with specifics on how each procedure should function may have influenced the overall distribution of procedure selection in this study.

Finally, how should problem size ideally be defined? At present, most contemporary research on simple arithmetic performance utilizes a dichotomy, irrespective of the operation being tested. For instance, addition problems are classified as small or large depending on the size of the sum, where a sum greater than 10 is considered large. For multiplication, any problem with a product of 25 or greater is classified as a large problem. For subtraction, any problem with a minuend above 10 is considered large. Take for example, the difference in operand magnitude when comparing subtraction and division problems. Using the inverse of addition problems, the largest subtraction problem in simple arithmetic is 18 – 9. In contrast, the corresponding division problem that uses the same operands (i.e., 18 ÷ 9) is only considered a small problem. If the increased attention paid to the operands is partially due to the size of the operands themselves, then comparing procedure use across operations based on eye movements may not be entirely accurate. Other researchers (Kirk & Ashcraft, 2001; Smith-Chant & LeFevre, 2003) have instead increased the number of levels by which to measure problem size (i.e., very small, small, large, and very large) in order to better understand when the problem-size effect shows the greatest impact. Adopting a similar definition might be more useful in determining where participants begin reallocating their attention.

Future Directions

Although the use of eye tracking remains largely underutilized in studies of arithmetic, the present findings highlight exciting new possibilities for future research. First, how might eye-pattern movements help us to better understand the procedures responsible for complex arithmetic tasks? Torbeyns et al. (2009) conducted an experiment in which young adult participants solved multi-digit subtraction problems that varied with respect to distance between minuend and subtrahend. For example, problems had operands relatively close in distance (e.g., 809 – 794), far in distance (e.g., 814 – 182) or somewhere in between (e.g., 820 – 460). Participants solved each problem under two no-choice procedure conditions. In one condition, participants had to solve each subtraction problem by means of direct subtraction (i.e., take the subtrahend away from the minuend to find the remainder). In the second condition, participants had to solve each subtraction problem by mean of indirect addition (i.e., what needs to be added to the subtrahend to reach the minuend). Direct subtraction was always slower than indirect addition, with the largest differences in performance observed for problems with a small distance between operands. Torbeyns et al. (2009) concluded that adults typically rely on their knowledge of addition when solving subtraction problems, especially when the difference in magnitude between operands is small. The use of eye tracking in an experiment such as this might help to augment the specific procedures participants utilize when solving more complex arithmetic problems.

Second, how might eye tracking help to augment traditional introspective methods used in studies that examine other domains of mathematical knowledge? Schneider et al. (2008) examined strategies used by elementary school children in a number line estimation task. Previous research involving this task postulated that young children choose different starting locations based on the magnitude of the target number. For instance, if the target number is between 1 and 20 (on a number line ranging from 1 to 100), children will begin estimating from the left of the number line. In contrast, if the number is closer to the right extreme, children begin estimation at the far right. Furthermore, children may use to the center to begin estimation. As predicted, Schneider et al. (2008) found that children in grades 1 through 3 made more fixations in the left, center, and right areas of the number line in comparison to the remainder of the number line, suggesting that eye movement patterns may validate the children's self-reports.

Finally, how might eye tracking contribute to our understanding of conceptual knowledge in mental arithmetic tasks? Participants also use a variety of procedures when they solve three-term problems (e.g., 5 + 2 – 2) that involve more than one operation. Dubé and Robinson (2010) and Robinson and Ninowski (2003) have suggested that adults are faster at solving these problems when they fully understand the concepts of inversion (i.e., a + b – b = a and no calculations are required) and associativity (i.e., that the problem 4 x 6 ÷ 2 does not need to be solved in a traditional left to right fashion). Although the above studies used self-reports to draw inferences about the relationship between different conceptual strategies and conventional measures of performance, eye tracking may help validate participants' descriptions.

Conclusions

Several explanations for the problem-size effect in simple arithmetic have been proposed over the past four decades. Recent findings and changes in experimental methodologies have supported the idea that the problem-size effect occurs, in part, because solver shift from using fast retrieval on small problems to using less-efficient procedures on large problems. Although participants' descriptions of their solution processes corroborate this position, the validity of self-reports themselves remains a topic of controversy in the field of mathematical cognition. In the wake of consistent challenges to the traditional introspective method of assessing cognitive processes involved in arithmetic problem solving, researchers have adopted alternative methods of assessing procedure use. One alternative is the use of eye-movement patterns to assess procedures and how they change as a function of problem-size. The present study utilized a combination of traditional performance measures and novel eye tracking measures to assess how the problem-size effect is distributed across problem components when adults solved basic division. Results suggest that different procedures rely on different problem components, especially for large division problems.

References

Ashcraft, M.H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In C.J. Brainerd, R. Kail, & J. Bisanz (Eds.), *Formal methods in developmental research* (pp, 302-338). New York: Springer-Verlag.

Ashcraft, M.H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 527-538.

Ashcraft, M.H., & Christy, K.S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1-6. *Journal for Research in Mathematics Education*, *26*, 296-421.

Barrouillet, P., & Thevenot, C. (2013). On the problem-size effect in small additions: Can we really discard any counting-based account? *Cognition*, *128*, 35-44.

Campbell, J.I.D., & Austin, S. (2002). Effects of response time deadlines on adults' strategy choices for simple addition. *Memory & Cognition*, *30*, 988-994.

Campbell, J. I. D., & Penner-Wilger, M. (2006). Calculation latency: The μ of memory and τ of transformation. *Memory & Cognition*, *34*, 217-226.

Campbell J.I.D., & Timm, J. (2000). Adults' strategy choices for simple addition: Effects of retrieval interference. *Psychonomic Bulletin & Review, 7,* 692-699.

Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130,* 299-315.

Campbell, J.I.D. (2008). Subtraction by addition. *Memory & Cognition*, *36*, 1094-1102.

Curtis, E. T., Huebner, M. G., & LeFevre, J-A. (2014). *The influence of problem size on fixation patterns for addition, subtraction, multiplication, and division*. Manuscript submitted for publication.

Dubé, A.K., & Robinson, K.M. (2010). The relationship between adults' conceptual understanding of inversion and associativity. *Canadian Journal of Experimental Psychology*, *64*, 60-66.

Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. *Cognition, 123*, 392-403.

Green, H.J., Lemaire, P., & Dufau, S. (2007). Eye movement correlates of younger and older adults' strategies for complex addition. *Acta Psychologica*, *125*, 257-278.

Heathcote, A., Popiel, S.J., & Mewhort, D.J.K. (1991). Analysis of response time distributions: An example using the stroop task. *Quantitative Methods in Psychology*, *109*, 340-347.

Hockley, W.E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 598-615.

Huebner, M.G., & LeFevre, J-A. (2014, July). *Use of procedural knowledge in mental subtraction: Evidence from eye movement patterns*. Poster presented at the 24th annual meeting of the *Canadian Society for Brain, Behaviour, and Cognitive Science*, Toronto, ON.

Huebner, M.G., MacKay, K.J., & LeFevre, J-A. (2013, November). *Individual differences and simple division: Use of eye tracking as a window on mental processing*. Poster presented at the 54th annual meeting of the *Psychonomic Society*, Toronto, ON.

Imbo, I., Vandierendonck, A., & Rosseel, Y. (2007). The influence of problem features and individual differences on strategic performance in simple arithmetic. *Memory & Cognition*, *35*, 454-463.

Imbo, I., & Vandierendonck, A. (2008). Practice effects on strategy selection and strategy efficiency in simple mental arithmetic. *Psychological Research*, *72*, 528-541.

Kirk, E. P., & Ashcraft, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 157-175.

Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology*, *4*, 35-45.

LeFevre, J.-A., Bisanz, J., Daley, K. E., Buffone, L., Graham, S. L., & Sadesky, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General, 125,* 284-306.

LeFevre, J-A., DeStefano, D., Penner-Wilger, M., & Daley, K.E. (2006). Selection of Procedures in Mental Subtraction. *Canadian Journal of Experimental Psychology*, *60*, 209-220.

LeFevre, J-A., & Liu, J. (1997). The role of experience in numerical skill: Multiplication performance in adults from Canada and China. *Mathematical Cognition*, *3*, 31-62.

LeFevre, J-A., & Morris, J. (1999). More on the relation between division and multiplication in simple arithmetic: Evidence for mediation of division solutions via multiplication. *Memory & Cognition, 27,* 803-812.

LeFevre, J-A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem-size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 216-230.

Leth-Steensen, C., Elbaz, Z.K., & Douglas, V.I. (2000). Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychologica*, *104*, 167-190.

Mauro, D.G., LeFevre, J-A., & Morris, J. (2003). Effects of problem format on division and multiplication performance: Division facts are mediated via multiplication-based representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 163-170.

McCloskey, M., Harley, W., & Sokol, S. (1991). Models of arithmetic fact retrieval: An evaluation in light of findings from normal and brain-damaged subjects. *Journal of Experimental Psychology*, *17*, 377-397.

Penner-Wilger, M., Leth-Steensen, C., & LeFevre, J-A. (2002). Decomposing the problem-size effect: A comparison of response time distributions across cultures. *Memory & Cognition*, *30*, 1160-1167.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446-461.

Ratcliff, R., & Murdock B.B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*, 190-214.

Robinson, K.M., Arbuthnott, K.D., & Gibbons, K.A. (2002). Adults' representations of division facts: A consequence of learning history? *Canadian Journal of Experimental Psychology*, *56*, 302-309.

Robinson, K.M., & Ninowski, J.E. (2003). Adults' understanding of inversion concepts: How does performance on addition and subtraction inversion problems compare to performance on multiplication and division inversion problems? *Canadian Journal of Experimental Psychology*, *57*, 321-330.

Robinson, K.M., Arbuthnott, K.D., Rose, D., McCarron, M.C., Globa, C.A., & Phonexay, S.D. (2006). Stability and change in children's division strategies. Journal of *Experimental Child Psychology*, *93*, 224-238.

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

Seyler, D.J., Kirk, E.P., & Ashcraft, M.H. (2003). Elementary subtraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1339-1352.

Siegler, R.S. (1989). Hazards of mental chronometry: An example from children's subtraction. *Journal of Educational Psychology*, *81*, 497-506.

Smith-Chant, B.L., & LeFevre. J-A. (2003). Doing as they are told and telling it like it is: Self-reports in mental arithmetic. *Memory & Cognition*, *31*, 515-528.

Sowinski, C., Dunbar, K., & LeFevre, J-A. *Calculation Fluency Test*. Technical Report, Math Cognition Laboratory, Carleton University.

Verschaffel, L., De Corte, E., Gielen, I., & Struyf, E. (1994). Clever rearrangement strategies in children's mental arithmetic: A confrontation of eye-movement data and verbal protocols. In J. E. H. V. Luit (Ed.), *Research on learning and instruction of mathematics in kindergarten and primary school* (pp. 153–180).

Doetincham, Netherlands: Graviant Publishing Company.

Zhou, F., Zhao, Q., Chen, C., & Zhou, X. (2012). Mental representations of arithmetic facts: Evidence from eye movement recordings supports the preferred operand-order-specific representation hypothesis. *Quarterly Journal of Experimental Psychology*, *65*, 661–74.