Fact Retrieval Deficits in Low Achieving Children and Children With Mathematical Learning Disability

David C. Geary, Mary K. Hoard and Drew H. Bailey

J Learn Disabil published online 20 January 2011
DOI: 10.1177/0022219410392046

Hammill Institute on Disabilities

Journal of Learning Disabilities
XX(X) 1–17

DOI: 10.1177/0022219410392046
http://journaloflearningdisabilities
.sagepub.com
SAGE

Fact Retrieval Deficits in Low Achieving Children and Children With Mathematical Learning Disability

David C. Geary,[1] Mary K. Hoard,[1] and Drew H. Bailey[1]

Abstract

Using 4 years of mathematics achievement scores, groups of typically achieving children (n = 101) and low achieving children with mild (LA–mild fact retrieval; n = 97) and severe (LA–severe fact retrieval; n = 18) fact retrieval deficits and mathematically learning disabled children (MLD; n = 15) were identified. Multilevel models contrasted developing retrieval competence from second to fourth grade with developing competence in executing arithmetic procedures, in fluency of processing quantities represented by Arabic numerals and sets of objects, and in representing quantity on a number line. The retrieval deficits of LA–severe fact retrieval children were at least as debilitating as those of the children with MLD and showed less across-grade improvement. The deficits were characterized by the retrieval of counting string associates while attempting to remember addition facts, suggesting poor inhibition of irrelevant information during the retrieval process. This suggests a very specific form of working memory deficit, one that is not captured by many typically used working memory tasks. Moreover, these deficits were not related to procedural competence or performance on the other mathematical tasks, nor were they related to verbal or nonverbal intelligence, reading ability, or speed of processing, nor would they be identifiable with standard untimed mathematics achievement tests.

Keywords

longitudinal study, learning disability, mathematics, mathematical cognition, memory retrieval, working memory, number line, number processing

Children who have difficulty learning mathematics are at risk for later underemployment and will experience obstacles meeting many common demands of the modern world (Every Child a Chance Trust, 2009; Rivera-Batiz, 1992). The research efforts devoted to identifying the cognitive mechanisms that contribute to these learning impediments have increased substantially over the past two decades and have yielded many insights (Berch & Mazzocco, 2007). Recent studies, for instance, indicate different patterns of cognitive deficit for children with severe and mild learning difficulties (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Jordan, Hanich, & Kaplan, 2003; Murphy, Mazzocco, Hanich, & Early, 2007). Children with severe difficulties, hereafter mathematically learning disabled (MLD), tend to score at or below the 10th national percentile on mathematics achievement tests grade after grade, perform poorly on many mathematical cognition tasks, and tend to have low-average scores in reading, working memory, and general intelligence (IQ). The children with less severe difficulties tend to have average reading ability, IQ, and working memory competencies but score between the 10th and 25th percentiles on mathematics achievement tests across grades (e.g., Murphy et al., 2007). These children appear to have more circumscribed mathematical cognition deficits and are described as low achieving (LA).

Despite progress, there are many unresolved issues, including debate regarding the specific nature of the mathematical deficits of children with MLD and LA children and the underlying mechanisms. The consistently identified mathematical deficits are difficulty learning basic arithmetic facts or retrieving them once they are learned, a developmental delay in the learning of arithmetical procedures, and poor comprehension of numeral magnitude (Butterworth, 2005; Butterworth & Reigosa, 2007; Geary, 1990, 1993; Jordan et al., 2003). Proposed mechanisms underlying these patterns range from a fundamental deficit in potentially inherent number and magnitude processing systems (Butterworth, 2005)

[1]University of Missouri, Columbia, MO, USA

Corresponding Author:
David C. Geary, University of Missouri, Department of Psychological Sciences, Columbia, MO 65211-2500
Email: GearyD@Missouri.edu

to domain-general deficits in working memory that make the learning of mathematics, and other academic domains, in school difficult (e.g., Swanson, 1993; Swanson & Sachse-Lee, 2001). A deficit in a fundamental magnitude representational system, as proposed by Butterworth (2005), could affect the learning of mathematical content that is dependent on an understanding of numerical magnitude and thus cascade into many or all of the observed mathematical deficits, including retrieval deficits. Alternatively, there may be several distinct deficits or delays underlying the mathematical performance of children with MLD and their LA peers.

Indeed, a recent brain imaging study in which children's fact retrieval was contrasted with their use of counting procedures to solve simple addition problems identified at least three brain regions that are engaged during the formation of long-term memories of basic facts and during the act of retrieval (Cho, Ryali, Geary, & Menon, 2010). These regions include areas of the prefrontal cortex that are important for controlled retrieval of information, the hippocampus that is critical for forming long-term memories, and areas of the parietal cortex involved in magnitude representations and more generally long-term memory storage (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Rivera, Reiss, Eckert, & Menon, 2005; Squire, Stark, & Clark, 2004). The implication is that there may be not only different forms of mathematical cognition deficits generally but also different forms of retrieval deficit in particular.

Accordingly, in a 3-year longitudinal assessment we compared and contrasted the fact retrieval deficits of children with MLD and LA children to determine if the developmental trajectory and nature (i.e., error patterns) of their deficits were similar or different. Following Jordan et al. (2003), we used a procedure that allowed us to separate our LA group into children with mild retrieval deficits, hereafter LA–mild fact retrieval, and children with severe retrieval deficits, LA–severe fact retrieval. We examined the retrieval characteristics (i.e., error rates and error patterns) of children in these groups and a group of typically achieving peers and whether they co-occurred with difficulties in other aspects of mathematical development, specifically competence in using arithmetical procedures, estimating magnitude on a mathematical number line, and fluency of processing small number sets. As elaborated below (Current Study), if a single fundamental numerical processing deficit is the source of the procedural, number processing, and retrieval deficits, as proposed by Butterworth (2005), then all of these deficits should co-occur and be only weakly related to reading ability and the domain-general learning abilities assessed by IQ, working memory, and processing speed measures. If the mechanisms contributing to the procedural, number processing, and retrieval deficits and delays are distinct, then some children should show one or two of these deficits but not all of them whereas other children should show a different pattern of deficits or delays. Finally, if the mechanisms underlying the fact retrieval deficits of children with MLD and LA–severe fact retrieval children differ, then different developmental and retrieval error patterns should emerge across these groups.

Mathematical Cognition

We provide basic reviews of the addition and number tasks used in this study and associated findings for groups of MLD and LA children.

Addition

The most thoroughly studied arithmetical competency is change in the distribution of strategies and procedures children use during problem solving (Siegler & Shrager, 1984). A common early strategy for solving simple addition problems is to count both addends. These counting procedures are sometimes executed by using fingers, the *finger counting strategy*, and sometimes without them, the *verbal counting strategy* (Siegler & Shrager, 1984). The min and sum procedures are two ways children count, whether or not they use their fingers (Groen & Parkman, 1972). The min procedure involves stating the larger-valued addend and then counting a number of times equal to the value of the smaller addend. The sum procedure involves counting both addends starting from 1; the less common max procedure involves stating the smaller addend and counting the larger one. Counting results in the development of long-term memory representations of basic facts, which then support the use of memory-based processes (Siegler & Shrager, 1984). The most common are *direct retrieval* of arithmetic facts and *decomposition* (6 + 7 might be solved by retrieving the answer to 6 + 6 and then adding 1 to this partial sum).

Relative to typically achieving children, children with MLD rely on finger counting for more years, adopt the min procedure at a later age, and commit more counting errors (Geary, 1993). The most consistent finding is that children with MLD and many LA children show a deficit in the ability to use retrieval-based processes (Barrouillet, Fayol, & Lathuliére, 1997; Jordan et al., 2003). These children correctly retrieve fewer facts and sometimes show a pattern of retrieval errors that involves intrusions of related but problem-irrelevant information into working memory (Passolunghi, Cornoldi, & De Liberto, 1999; Passolunghi & Siegel, 2004; also see Engle, Tuholski, Laughlin, & Conway, 1999). For the simple addition problems used in the current study, counting string associates of the addends appear to be the most common intrusions (Geary, Hamson, & Hoard, 2000). For 6 + 3, an answer of 7 or 4 would be coded as a counting sting intrusion because 7 follows 6 and 4 follows 3 in the counting string.

Number Line Representations

The number line is important from an educational perspective because individual differences in children's knowledge of the linear, mathematical number line are correlated with mathematics achievement (Booth & Siegler, 2006; Siegler & Booth, 2004). It is important from a cognitive perspective because children's learning of the number line may be based on a potentially inherent system that represents approximate magnitudes of quantities larger than 3 or 4 (e.g., Kadosh et al., 2007). This same system appears to be a component of the number processing module that Butterworth (2005) hypothesized may underlie many of the deficits in children with MLD and LA children.

The representational system engaged by this task is inferred from how the child makes placements on a physical number line. Placements based on use of the inherent number magnitude system result in a pattern that conforms to the natural logarithm (Ln) of the number (Feigenson, Dehaene, & Spelke, 2004; Gallistel & Gelman, 1992; Siegler & Opfer, 2003). In other words, use of this natural representational system results in the compression or spatial "crowding" of the values of larger magnitudes, such that the perceived difference between 91 and 94, for example, is psychologically smaller than the perceived distance between 1 and 4. With schooling, typically achieving children's number line placements gradually conform to the linear mathematical system (Siegler & Booth, 2004); the difference between two consecutive numbers is identical regardless of position on the line.

Geary et al. (2007; Geary, Hoard, Nugent, & Byrd-Craven, 2008) compared the number line performance of MLD, LA (mild and severe combined), and typically achieving first and second graders. Group differences emerged using group medians, with trial by trail assessments of whether the placement was consistent with a log or linear representation and with several measures of absolute error. The latter included the absolute difference between the child's placement and the correct placement independent of the representation guiding the placement and absolute degree of error for linear and log trials (e.g., degree to which the placement differed from the predicted log placement for log trials). The results confirmed that typically achieving children quickly learn the linear mathematical number line.

Children with MLD were more heavily dependent on the natural representational system, and consistent with Butterworth's (2005) hypothesis their representation of magnitude appeared to be more compressed than that of LA and typically achieving children. In first grade, the children with MLD did not appear to discriminate the magnitudes of smaller numerals as easily as the other children. This was the case even when these children used the natural representational system, which should make discriminations of smaller values (e.g., 1 vs. 4) easier than would use of the linear, mathematical representation. By the end of second grade, their performance was comparable to that of typically achieving children at the beginning of first grade, suggesting roughly a 2-year delay in the maturation of the magnitude representational system or in the construction of the linear representation from this system; the performance of the LA children suggested about a 1-year delay.

Rousselle and Noël (2007), in contrast, found no differences across MLD (defined as scores below the 15th percentile on a mathematics composite test) and typically achieving second graders on tasks that involved comparing collections of items beyond the subitizing range (below) but did find differences, favoring the typically achieving children, for speed and accuracy of comparing magnitudes represented by Arabic numerals. They argued the deficit of the children with MLD was the result of poor mapping of symbols (e.g., *5*) onto the underlying magnitude and not a deficit in the magnitude representational system per se. These results are not necessarily discrepant with Geary et al.'s (2008) findings, if the magnitude representational system simply matures more slowly in children with MLD; the second grader in Rousselle and Noël's study may have been too old to detect differences in maturational timing.

Number Sets

Subitizing represents another potentially inherent number-processing system, specifically for quickly apprehending the quantity of sets of one to four objects without counting (e.g., Starkey & Cooper, 1980). Children with MLD have a potential deficit in this system (Butterworth & Reigosa, 2007; Geary, Bailey, & Hoard, 2009; Geary et al., 2007; Koontz & Berch, 1996). Koontz and Berch (1996) assessed the ability of third and fourth graders with MLD and their typically achieving peers to apprehend, without counting, the quantity of small sets of items or corresponding Arabic numerals; for example, the children were asked to determine if combinations of Arabic numerals and number sets were the same (e.g., 2-■■) or different (e.g., 3-■■). Reaction time (RT) patterns for the typically achieving children indicated fast access to representations of quantities of two and three, regardless of whether the code was an Arabic numeral or number set. The children with MLD showed fast access to numerosity representations for the quantity of two but appeared to rely on counting to determine quantities of three. The results suggest that some children with MLD might not have an inherent nonverbal representation for numerosities of three or more; likely the representational system for three does not allow for reliable discrimination of two from three, in keeping with Butterworth's (2005) hypothesis.

Geary et al. (2007) developed the *Number Sets Test* (described below) to assess the fluency with which children process and add sets of objects and Arabic numerals to

match a target number, for example, whether the combination ●● 3 matches the target of 5. The items are similar to those used by Koontz and Berch (1996), albeit some involve magnitudes up to 9. Fluency should be aided by rapid subitizing, rapid access to the magnitudes represented by small Arabic numerals, and the ability to add and compare and contrast these magnitudes. Using a receiver operating characteristic (ROC) analysis, a sensitivity measure (d-prime) can be obtained by subtracting each participant's *z* score for misses (e.g., identifying ●●● 3 as equal to 5) from his or her *z* score for hits (MacMillan, 2002). This value provides a measure of sensitivity to number combinations that match the target value while controlling for the child's response bias (i.e., tendency to circle items whether or not they match the target).

The d-prime scores, but not response bias scores, are correlated with mathematics but not reading achievement, above and beyond the influence of IQ, working memory, and prior grade mathematics and reading achievement scores (Geary, Bailey, & Hoard, 2009). Children identified as MLD in third grade—as determined by achievement scores less than the 15th national percentile ranking—have lower d-prime scores in first grade. In fact, the first grade d-prime score was a better predictor of third grade MLD status than were first grade mathematics achievement scores. The test appears to tap very basic number processing competencies, and children with MLD have deficits in these competencies.

Current Study

The current study included 231 children classified into MLD, LA–mild fact retrieval, LA–severe fact retrieval, and typically achieving groups based on mathematics achievement scores from kindergarten to third grade (see the method section) and based on a forced addition fact retrieval task administered from second to fourth grade (Jordan et al., 2003). To recap, our goals are to determine if the fact retrieval deficits of LA–mild fact retrieval and LA–severe fact retrieval children are similar to those of children with MLD and to determine if these deficits covary with or are distinct from competence in executing arithmetical procedures and the number representation and processing competencies assessed by the number line and number sets measures. Inclusion of these number tasks allows us to test the hypothesis that a deficit in the ability to represent and process numerical magnitude underlies many of the other deficits found with children with MLD and their LA peers (Butterworth, 2005). Because performance on most cognitive tasks will be influenced by domain-general abilities, these are potential alternative sources of group differences on the mathematical cognition tasks. As a control, we conducted analyses of group differences on key mathematical cognition variables, after controlling for these potential influences, that is, IQ, working memory, and processing speed as well as reading ability (Fuchs, Geary, Compton, Fuchs, & Hamlett, 2010).

Method

Participants

The data are from a prospective study of mathematical development and disabilities (Geary, 2010; Geary et al., 2007). For the original sample, all kindergarten children from 12 elementary schools in the same district were invited to participate. The schools serve children from a wide range of socioeconomic backgrounds, and several of the schools have had a high proportion of children with MLD or LA children in previous studies. Parental consent and child assent were received for 37% (n = 311) of these children, and 305 of them completed the first wave of testing. The mathematics curriculum when the children started the study was *Investigations in Number, Data, and Space* (Scott Foresman, 1999), and they continued with this curriculum throughout the grades analyzed here.

Nearly complete achievement and cognitive data were available through fourth grade for 257 of these children. Comparison of these children to the 48 children who dropped from the study after the first assessment revealed trends for group differences in kindergarten IQ scores (M = 97, SD = 18; M = 101, SD = 14; for dropped and retained, respectively), $F(1, 303) = 3.1$, $p = .08$, and the percentile rank on the reading achievement test (M = 65, SD = 27; M = 72, SD = 22), $F(1, 303) = 3.44$, $p = .06$, but not for the percentile rank on the mathematics achievement test (M = 47, SD = 28; M = 53, SD = 25), $F(1, 303) = 2.62$, $p > .10$. We do not have socioeconomic status for the children's parents, but we do have the percentage of students receiving free or reduced-price lunches in each of the schools they attended. There were across-school differences in the numbers of children who dropped or stayed in the study, $\chi^2(12) = 31.47$, $p < .002$. Children who dropped attended schools with a higher percentage of students receiving free or reduced-price lunches (43%) than children who stayed in the study (35%), $F(1, 303) = 5.81$, $p < .02$.

To identify stable groups with different achievement trajectories, Geary, Bailey, Littlefield, et al. (2009) used latent class growth trajectory analyses applied to kindergarten through third grade mathematics achievement scores, inclusive. One advantage of this approach is that it makes achievement cutoff scores unnecessary because it clusters together children with similar achievement start points and grade to grade change in achievement scores. We used these clusters but dropped children with IQ scores less than 85 and a high achieving group of children, leaving 101 (52 males), 115 (39 males), and 15 (10 males) children, respectively, for the typically achieving, LA, and MLD groups.

Following Jordan et al. (2003), we used a forced retrieval of addition facts task to determine the extent of the retrieval deficits of children with MLD and to explore the extent to which LA children also showed these deficits. Based on performance on one aspect of this task, the latter group was divided into children with mild retrieval (LA–mild fact retrieval; $n = 97$, 35 males) and more severe (LA–severe fact retrieval; $n = 18$, 4 males) deficits: When instructed to use only retrieval to solve simple addition problems, children who correctly retrieved answers to fewer than 6 of 14 simple problems in second, third, *and* fourth grade were classified as LA–severe fact retrieval. When this criterion was applied across all grades, it resulted in the identification of children who could correctly retrieve only two to three answers for the 14 problems, with no across-grade improvement, $F(2, 30) < 1$.

At the time of the fall, second grade assessment, the mean across-group ages ranged between 85 and 87 months ($p > .50$). There were more girls in both the LA–mild fact retrieval and LA–severe fact retrieval groups; $\chi^2 = 7.52$, 5.56, ps $< .02$. The ethnic distribution was 72% White and 10% Black, and most of the remaining children were Asian, Hispanic, or of mixed race. Children with MLD came from schools with a higher percentage of students receiving free or reduced-price lunches (45%, $p < .05$) than did children in the remaining groups (range from 31% to 36%); the latter groups did not differ from one another (ps $> .05$).

Standardized Measures

Intelligence. The children were administered the *Raven's Coloured Progressive Matrices* (Raven, Court, & Raven, 1993) in kindergarten and the Vocabulary and Matrix Reasoning subtests of the *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) in first grade. The raw *Raven's* scores were standardized ($M = 100$, $SD = 15$) using the 305 children in our sample who completed the test.

Achievement. The children were administered the Numerical Operations and Word Reading subtests from the *Wechsler Individual Achievement Test–II–Abbreviated* (Wechsler, 2001). The former assesses number discrimination, rote counting, number production, basic addition and subtraction, multidigit addition and subtraction, and some multiplication and division. The latter includes matching and identifying letters, rhyming, beginning and ending sounds, phoneme blending, and word recognition.

Mathematical Tasks

We assessed retrieval with two addition tasks; for one, children choose any strategy to solve the presented problems (choice task), and for the other they are asked to only use retrieval (forced retrieval task). To provide a contrast, we analyzed performance on two other mathematical tasks, the *Number Sets Test* and a number line estimation task.

Choice task. One at a time, 14 addition problems were horizontally presented at the center of a computer monitor (i.e., 3 + 6, 5 + 3, 7 + 6, 3 + 5, 8 + 4, 2 + 8, 9 + 7, 2 + 4, 9 + 5, 7 + 2, 9 + 8, 4 + 7, 2 + 5, and 3 + 9). The problems consisted of the integers 2 through 9, with the constraint that the same two integers (e.g., 2 + 2) were never used in the same problem; half of the problems summed to 10 or less, and the smaller valued addend appeared in the first position for half of the problems. The child was asked to solve each problem (without paper and pencil) as quickly as possible without making too many mistakes. It was emphasized that the child could use whatever strategy was easiest to get the answer and was instructed to speak the answer into a microphone that was interfaced with the computer, which in turn recorded RT from onset of problem presentation to activation of the microphone.

After solving each problem the child was asked to describe how he or she got the answer. Based on the child's description and the experimenter's observations, the trial was classified based on problem-solving strategy; the four most common were counting fingers, verbal counting, retrieval, and decomposition. Counting trials were further classified as min, sum, max, or other (see Geary et al., 2007). The combination of experimenter observation and child reports immediately after each problem is solved has proven to be a useful measure of children's strategy choices (Geary, 1990; Siegler, 1987). The usefulness of this information is supported by previous findings showing that finger counting trials have the longest RTs, followed by verbal counting, decomposition, and direct retrieval. The same pattern emerged with the current study. As an example, for second grade, mean RTs for correct retrieval trials were 2,789 (SD = 1,892) ms, and those for counting fingers trials were 6,662 (SD = 4,153) ms; means for verbal counting (M = 4,980, SD = 3,928) and decomposition (M = 4152, SD = 2784) were in between these. All pairwise comparisons of correct mean RTs were significant ($|t|$s > 7.49, ps $< .05$).

Forced addition fact retrieval. Beginning in second grade, the forced addition fact retrieval problems were administered after the choice problems and were the reverse of the 14 original problems (e.g., 6 + 3, 3 + 6). The only change was that the children were instructed to solve the new problems using only retrieval. They were instructed to try to remember the answer as quickly as they could and were instructed not to count or use any other type of strategy; if they could not remember, they were told that it was okay to guess. This task has been used in previous research to assess the extent of mastery of basic facts and retrieval characteristics, including intrusion errors (Geary et al., 2000; Jordan et al., 2003; Jordan & Montani, 1997; Russell & Ginsburg, 1984; Siegler & Shrager, 1984).

Number sets. The test assesses the speed and accuracy with which children understand and manipulate small, exact numerosities within the subitizing range, and quantities just

beyond this range (< 10) while transcoding between those quantities and their Arabic numerals. Two types of stimuli are used: objects (e.g., squares) in a 0.5 in. square and an Arabic numeral (18-point font) in a 0.5 in. square. Stimuli are joined in domino-like rectangles with different combinations of objects and numerals. These dominos are presented in lines of five across a page. The last two lines of the page show three 3-square dominos. Target sums (5 or 9) are shown in larger font at the top the page. On each page, 18 items match the target, 12 are larger than the target, 6 are smaller than the target, and 6 contain 0 or an empty square.

The tester begins by explaining two items matching a target sum of 4, then uses the target sum of 3 for practice. The measure is then administered. The child is told to move across each line of the page from left to right without skipping any, to "circle any groups that can be put together to make the top number, 5 (9)," and to "work as fast as you can without making many mistakes." The child has 60 s per page for the target 5, 90 s per page for the target 9. Time limits were chosen to avoid ceiling effects and to assess fluent recognition and manipulation of quantities.

Geary et al. (2007) found that first graders' performance was consistent across target number and item content (e.g., whether the rectangle included Arabic numerals or shapes) and could thus be combined to create an overall frequency of hits ($\alpha = .88$), correct rejections ($\alpha = .85$), misses ($\alpha = .70$), and false alarms ($\alpha = .90$). As described earlier, the d-prime measure derived from ROC analyses captures variance unique to mathematics achievement after controlling for domain-general cognitive competencies and was thus used in the current analyses.

Number line estimation. A series of twenty-four 25 cm number lines containing a blank line with two endpoints (0 and 100) was presented, one at a time, to the child with a target number (e.g., 45) in a large font printed above the line. The child's task was to mark on the line where the target number should lie (for a detailed description, see Siegler & Booth, 2004). Siegler and Opfer (2003) used group-level median placements fitted to linear and log models to make inferences about the modal representation children were using to make the placements, and for individual difference analyses they used an accuracy measure. Accuracy is defined as the absolute difference between the child's placement and the correct position of the number. For the number 45, placements of 35 and 55 produce difference scores of 10.

Other potential individual differences measures include the frequency with which children make placements consistent with a linear representation of the line or placements that conform to the natural log of the numbers, suggesting use of the approximate magnitude representational system. To determine the best measure of children's understanding of the linear number line, we correlated (using data from all available children) absolute number line error, the percentage of trails consistent with use of a linear representation and the degree of error for these trails, and the percentage of trails consistent with use of a log representation and the degree of error for these trails (i.e., the degree to which the placement differed from the predicted log placement; see Geary et al., 2007; Geary et al., 2008) with Numerical Operations achievement scores in first and second grade. We included first grade because use of the log representation becomes less common in later grades, and thus the influence of this strategy on number line performance becomes difficult to assess (Geary et al., 2008; Siegler & Booth, 2004). The best single predictor of Numerical Operations scores was absolute number line error in first, $r(287) = -.46$, and second, $r(269) = -.45$, grades.

We then simultaneously regressed absolute error on the percentage of linear and log trials and corresponding error rates to determine the sources of the absolute error scores. In first grade, the degree of absolute error increased with increases in the percentage of log trials, $\beta = .40$, $t(278) = 7.45$, degree of error on log trials, $\beta = .64$, $t(278) = 40.05$, and degree of error on linear trials, $\beta = .17$, $t(278) = 11.09$, $R^2 = .96$. In second grade, the degree of absolute error increased with increases in the percentage of log trials, $\beta = .58$, $t(200) = 3.77$, and degree of error on log trials, $\beta = .48$, $t(200) = 4.56$, $R^2 = .40$. The results suggest that (a) the absolute error score is a good predictor of mathematics achievement scores, as found by Siegler and colleagues (Booth & Siegler, 2006; Siegler & Booth, 2004), and (b) the magnitude of these errors is related to the frequency with which children use the log strategy and the extent to which these placements differ from the predict log placement. On the basis of these analyses, we chose the absolute difference variable as the measure of the extent to which children have learned the linear, mathematical number line. The overall score is the mean of these differences across trials.

Working Memory and Speed of Processing

Working memory. The *Working Memory Test Battery for Children* (WMTB-C; Pickering & Gathercole, 2001) consists of nine subtests that assess the central executive, phonological loop, and visuospatial sketch pad. All of the subtests have six items at each span level. Across subtests, the span levels range from one to six to one to nine. Passing four items at one level moves the child to the next. At each span level, the number of items (e.g., words) to be remembered is increased by one. Failing three items at one span level terminates the subtest. The subtests are used to define measures of three core working memory systems, the central executive, phonological loop, and visuospatial sketch pad. Our sample of 270 first graders (below) is larger than the standardization sample at this age, and thus we standardized ($M = 100$, $SD = 15$) the raw scores from our sample and used these in the analyses.

Central executive. The central executive is assessed using three dual-task subtests. Listening Recall requires the child to determine if a sentence is true or false and then recall the last word in a series of sentences. Counting Recall requires the child to count a set of 4, 5, 6, or 7 dots on a card and then recall the number of counted dots at the end of a series of cards. Backward Digit Recall is a standard format backward digit span.

Phonological loop. Digit Recall, Word List Recall, and Nonword List Recall are standard span tasks with differing content stimuli; the child's task is to repeat words spoken by the experimenter in the same order as presented by the experimenter. In the Word List Matching task, a series of words, beginning with two words and adding one word at each successive level, is presented to the child. The same words, but possibly in a different order, are then presented again, and the child's task is to determine if the second list is in the same or different order than the first list.

Visuospatial sketch pad. Block Recall is another span task, but the stimuli consist of a board with nine raised blocks in what appears to the child as a "random" arrangement. The blocks have numbers on one side that can be seen only from the experimenter's perspective. The experimenter taps a block (or series of blocks), and the child's task is to duplicate the tapping in the same order as presented by the experimenter. In the Mazes Memory task, the child is presented a maze with more than one solution and a picture of an identical maze with a path drawn for one solution. The picture is removed, and the child's task is to duplicate the path in the response booklet. At each level, the mazes get larger by one wall.

Speed of processing. Using the same stimuli as Mazzocco and Myers (2003), two Rapid Automatized Naming (RAN; Denckla & Rudel, 1976) tasks were used to assess processing speed. The child is presented with 5 letters or numbers to first determine if the child can read the stimuli correctly. After these practice items, the child is presented with a 5 × 10 matrix of 50 incidences of these same letters or numbers and is asked to name them as quickly as possible without making any mistakes. RT is measured via a stopwatch, and errors and reversals for the letters b and d and p and q are recorded. For each type of stimulus (letters or numbers), the task generates an RT, number correct, and number of reversals for letters. Errors and reversals were too infrequent for meaningful analysis, and thus only RTs were used.

Procedure

All children were tested in the spring of their kindergarten year and in the fall and spring of all successive grades. The spring assessments included the achievement and intelligence measures and the number line task beginning in second grade. Fall testing included the remaining mathematical tasks and the RAN tasks. The majority of children were tested in a quiet location at their school site and occasionally on the university campus or in a mobile testing van. Testing on campus or in the van occurred for children who had moved out of the school district. Campus testing occurred for a few children whose families returned to Columbia to visit family, but most of these children were tested in the van, which is equipped with a testing table, backup battery, camera to videotape the sessions, and other modifications needed for the assessments. The van allows us reduce attrition by retaining children who move within a 5-hr drive (one way) of the university in the study.

The WMTB-C was added to the study in the summer following kindergarten, and we received parental permission to assess 270 children from the original sample; this included 97 of the 101 typically achieving children, 90 of the 97 LA–mild fact retrieval children, 18 of the 18 LA–severe fact retrieval children, and 13 of the 15 children with MLD. For the majority of children, the battery was administered in the testing van during first grade. The assessment required about 60 min and occurred when the child was not in school (e.g., weekend).

Results

In the first section, group differences and across-grade changes were analyzed using multilevel modeling; specifically, PROC MIXED (SAS Institute, 2004). Group (i.e., typically achieving, LA–mild fact retrieval, LA–severe fact retrieval, and MLD) and grade were estimated as fixed effects and individual slopes and intercepts as random effects. Effectively this means that we ran a separate regression for each group to assess how the groups' slopes and intercepts differed, and then we ran a separate regression for each participant to see how participants' slopes and intercepts differed from each other within their group. The corresponding intercept values estimate the mean group differences in second grade, and the slope values estimate the rate of change from second (coded 0) to third (coded 1) to fourth (coded 2) grades. Significant intercept or slope effects are followed by comparisons of pairwise means across groups.

In the second section, we examined whether group differences on key mathematical cognition tasks identified in the first section could be related to the domain-general abilities represented by IQ, working memory, and processing speed as well as reading ability.

Multilevel Models

Standardized measures. Standardized scores with a mean of 50 (SD = 10) are provided by the WASI (Wechsler, 1999) for the Vocabulary and Matrix Reasoning subtests, but we rescaled these to a mean of 100 (SD = 15) to make the scores comparable to those from the *Raven's* test. Repeated

Table 1. Intelligence and Achievement Test Scores

		IQ				Second Grade				Third Grade				Fourth Grade			
		Nonverbal		Verbal		Word Reading		Numerical Operations		Word Reading		Numerical Operations		Word Reading		Numerical Operations	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
TA	101	104^{a}	12	106^{a}	12	74^{a}	20	55^{a}	22	66^{a}	20	55^{a}	22	66^{a}	20	54^{a}	24
LA-M	97	98^{b}	10	97^{b}	9	53^{b}	27	21^{b}	12	49^{b}	23	20^{b}	14	49^{b}	23	27^{b}	21
LA-S	18	101^{ab}	15	100^{ab}	12	57^{b}	29	17^{bc}	12	53^{b}	23	17^{b}	12	52^{b}	22	24^{b}	15
MLD	15	91^{b}	7	82^{c}	5	11^{c}	9	4^{c}	3	13^{c}	11	3^{c}	2	21^{c}	13	3^{c}	2

Note: TA = typically achieving; LA-M = low achieving with mild retrieval deficits; LA-S = low achieving with counting fingers min procedure severe retrieval deficits; MLD = mathematical learning disability. Group means in the same column with different superscripts differ significantly ($p < .05$). Significance values for achievement scores control for group differences in IQ; $p = .062$ for the third grade Word Reading and Numerical Operations contrast of the LA-S and MLD groups.

measures analyses of variance were then run separately for each group, with the three IQ scores as within-subjects variables. The only significant effect emerged for children with MLD, $F(2, 28) = 4.74$, $p < .02$. Follow-up analyses revealed lower Vocabulary, hereafter verbal IQ, scores than Matrix Reasoning, $F(1, 14) = 4.43$, $p = .054$, and *Raven's*, $F(1, 14) = 5.61$, $p < .05$, scores. The two latter scores did not differ ($p > .15$), and thus their mean was used as a measure of nonverbal IQ. Mean verbal and nonverbal IQ scores are shown for all groups in Table 1. Group differences were significant for verbal, $F(3, 227) = 20.48$, $p < .0001$, $d = 1.60$, and nonverbal, $F(3, 227) = 8.48$, $p < .0001$, $d = 0.87$, IQ (see Note 1). Follow-up honestly significant difference tests revealed the children with MLD had lower verbal IQ scores than all other groups (ps < .05) and the LA–mild fact retrieval group had a lower score than the typically achieving children ($p < .01$). There were only two significant group differences for nonverbal IQ; the typically achieving children had higher scores than children in the LA–mild fact retrieval and MLD groups (ps < .05).

Achievement. Covarying IQ scores, the groups differed in national percentile ranking for both achievement tests at each grade level (ps < .001); the typically achieving group had higher percentile rankings than the LA–mild fact retrieval and LA–severe fact retrieval groups, who did not differ from each other but who scored higher than the MLD group, ds (based on least square means) = 1.56 to 1.87.

The multilevel results for raw achievement scores confirmed significant group, $F(3, 221) = 118.17, 31.72$, grade, $F(1, 227) = 420.41, 845.41$, and interaction effects, $F(3, 221) = 4.48, 3.07$, for the Numerical Operations and Word Reading tests, respectively (ps < .005, except $p < .03$ for the Word Reading interaction). For both tests, pairwise comparisons indicated the typically achieving group had a higher intercept than all other groups ($p < .005$), the LA–mild fact retrieval and LA–severe fact retrieval groups did not differ (ps > .25), but both had higher intercepts than the MLD group (ps < .01). All groups showed raw score improvements on both tests. None of the pairwise second to fourth grade slopes differed for Word Reading (ps > .10) but did for Numerical Operations; the slope for the typically achieving group was higher than that of the MLD (ps > .001) and the LA–mild fact retrieval ($p = .059$) groups, but not the LA–severe fact retrieval group ($p > .50$). The two LA groups did not differ ($p > .25$), but both had steeper slopes than the MLD group (ps < .05).

The bottom line is that the typically achieving, LA–mild fact retrieval, and LA–severe fact retrieval groups had nonverbal and verbal IQ scores in the average range and the children with MLD had low-average nonverbal scores and below average verbal scores. In second grade, the typically achieving children had higher raw reading and mathematics achievement scores than children in all other groups, the two LA groups did not differ, but both outscored the children with MLD. The rate of second to fourth grade growth in word reading fluency did not differ across groups, but the mathematical knowledge of the typically achieving children grew faster than that of the children in the MLD and LA–mild fact retrieval groups. The two LA groups did not differ in rate of growth in mathematical knowledge, but both grew faster than that of the MLD group.

Addition strategy choice. The mix of strategies used to solve the choice problems is shown in Table 2. The intercepts represent group means for the percentage of problems on which the strategy was used in second grade, and the slope represents the change in percentage usage across grades. For example, the typically achieving children used finger counting to solve 17% of the addition problems, on average, in second grade but used this strategy less frequently across grades; the average percentage usage of finger counting for this group in fourth grade was 4 (i.e., $17 - (2 \times 6.5)$).

The main effect for group was significant for each of the four strategies, $F(1, 227) = 34.99, 6.63, 23.15, 24.31$, ps < .02, for counting fingers, verbal counting, retrieval, and decomposition, respectively. The main effect for slope was also significant for each strategy, $F(3, 227) = 14.29$,

Table 2. Group Differences in Intercept and Slope for Percent Usage and Errors for Addition Strategy Choices

	Strategy											
	Finger Counting		Verbal Counting		Retrieval		Decomposition		CF Min		VC Min	
	Int.	Slope	Int.	Slope	Int.	Slope	Int.	Slope	Int.	Slope	Int.	Slope
Percent usage												
TA	17^{a} (2.5)	−6.5^{a} (5.0)	15^{ac} (2.1)	−4.4^{a} (1.2)	37^{a} (2.0)	5.7^{a} (1.2)	28^{a} (1.8)	5.4^{ab} (1.1)	96^{a} (2.3)	1.0^{a} (1.5)	96^{a} (7.3)	1.6^{a} (1.3)
LA-M	36^{b} (2.5)	−11.0^{a} (5.0)	25^{bd} (2.1)	−5.5^{a} (1.2)	25^{b} (2.1)	8.3^{a} (1.2)	13^{b} (1.9)	8.0^{a} (1.1)	93^{a} (1.9)	2.2^{a} (1.1)	96^{a} (7.3)	0.7^{a} (1.1)
LA-S	40^{b} (5.9)	−6.9^{a} (5.8)	35^{b} (5.0)	−4.7^{a} (2.8)	13^{b} (6.9)	5.8^{a} (2.8)	7^{bc} (4.4)	5.6^{ab} (2.6)	96^{a} (4.4)	1.4^{a} (2.4)	92^{a} (8.0)	4.4^{a} (2.3)
MLD	48^{b} (6.4)	−5.2^{a} (3.4)	21^{cd} (5.4)	3.2^{b} (3.0)	22^{b} (5.3)	2.1^{a} (3.1)	0^{c} (4.7)	1.8^{b} (2.9)	61^{b} (4.9)	15.5^{b} (2.6)	85^{b} (5.0)	6.2^{a} (3.0)
Percent error												
TA	6^{a} (1.9)	−3.0^{ab} (1.7)	4^{a} (2.0)	−1.5^{a} (1.5)	3^{a} (1.5)	−0.8^{a} (0.9)	2 (13.2)	−0.5 (5.0)	—	—	—	—
LA-M	7^{a} (1.5)	−0.4^{b} (1.2)	8^{a} (1.8)	−2.9^{a} (1.2)	5^{a} (1.6)	−1.5^{a} (0.9)	6 (13.2)	−2.3 (5.0)	—	—	—	—
LA-S	8^{a} (3.6)	2.3^{b} (2.4)	7^{a} (4.1)	3.4^{b} (2.5)	15^{b} (4.3)	−5.4^{a} (2.5)	5 (11.5)	3.3 (3.5)	—	—	—	—
MLD	20^{b} (4.0)	−6.6^{a} (2.6)	31^{b} (5.4)	−10.0^{c} (3.2)	59^{c} (4.7)	−20.8^{b} (2.8)	0 (9.3)	3.3 (3.5)	—	—	—	—

Note: CF min = counting fingers min procedure; VC min = verbal counting min procedure; TA = typically achieving; LA-M = low achieving with mild retrieval deficits; LA-S = low achieving with severe retrieval deficits; MLD = mathematical learning disability. Group means in the same column with different superscripts differ significantly ($p < .05$). The absolute frequency of decomposition errors was too low to contrast the group means for slope and intercept; group differences were not assessed for percent error in the Decomposition columns for Int. and Slope.

6.62, 10.84, 20.60, ps < .0003, for counting fingers, verbal counting, retrieval, and decomposition, respectively. However, none of the interactions were significant (ps > .05). Follow-up contrasts confirmed the typically achieving children used finger counting less frequently and retrieval more frequently in second grade than did children in the other groups ($|t|$s > 2.7, ps < .01), who did not differ from one another (ps > .05); children in all groups showed similar across-grade decreases in finger counting and increases in retrieval. The results were more complex for verbal counting and decomposition, but in general typically achieving second graders used verbal counting less frequently (although the contrast with the MLD group was not significant) and decomposition more frequently than did children in the other groups, who used verbal counting in roughly similar frequencies. The children with MLD almost never used decomposition in second grade, and the children in both LA groups used it to solve 1 or 2 of the 14 problems, on average. Across grades, typically achieving children and children in both LA groups used verbal counting less and decomposition more frequently; the group differences for the corresponding slopes were not significant ($|t|$s < 1.7, ps > .10). The children with MLD used verbal counting more frequently across grades, and a few of them occasionally used decomposition in third and fourth grade.

For use of the min procedure, the slope effect was significant for both finger, $F(1, 129) = 25.00$, and verbal, $F(1, 53) = 9.92$, counting (ps < .005). The group effect, $F(3, 73) = 14.72$, and the interaction, $F(3, 73) = 8.58$ (ps < .001), were significant for finger counting, but neither was significant for verbal counting (ps > .15). Basically, in comparison to children in the other groups, the children with MLD used the min procedure while finger counting less often in second grade but caught up by fourth grade. Second graders with MLD made at least 2.5 times more counting errors, whether using their fingers or not, as children in the other groups ($|t|$s > 2.25, ps < .05); the children in the latter groups did not differ ($|t|$s < 1). The LA–severe fact retrieval children committed more errors than the typically achieving and LA–mild fact retrieval children ($|t|$s > 2.17, ps < .05), and the children with MLD committed more errors than any other group ($|t|$s > 6.48, ps < .001). The across-grade changes indicated improvements in counting accuracy and fewer retrieval errors for the children with MLD but a more mixed pattern for the LA–severe fact retrieval children.

The gist is that in second grade the children in the MLD and two LA groups relied more heavily on finger and verbal counting than the typically achieving children, who in turn relied more heavily on retrieval and decomposition. Although all groups showed a second to fourth grade shift to less frequent use of finger counting and more frequent use of retrieval and decomposition, the group differences in strategy usage remained through fourth grade. Confirming previous findings, the children with MLD showed a grade-related increase in verbal counting in lieu of finger counting and an increase in use of the min procedure. Also in keeping with previous findings, the second graders with MLD committed more counting and retrieval errors than did the children in the other groups (Geary, 1993), but they narrowed the gap by fourth grade. With respect to the goals of the current study, the core findings were the low rates of retrieval errors for typically achieving and LA–mild fact retrieval groups and the higher but still modest errors rates of the children in the LA–severe fact retrieval group.

Forced addition fact retrieval. As per instructions and based on the child's report and experimenter observation, between

Table 3. Group Differences in Intercept and Slope for Percent Retrieval and Intrusion Errors for No-Choice Problems and Procedural Competence for Choice Problems

	Percentage of Retrieval Errors				Percentage of Retrieval Errors That Were Intrusions				Procedural Competence			
	Int.		Slope		Int.		Slope		Int.		Slope	
TA	37[a]	(2.5)	−1.9[a]	(1.5)	4[a]	(1.0)	−0.2[a]	(0.6)	24[a]	(0.7)	1.2[a]	(0.5)
LA-M	55[b]	(2.6)	−8.9[b]	(1.5)	8[b]	(1.0)	−1.9[a]	(0.6)	24[a]	(0.6)	0.6[a]	(0.4)
LA-S	84[c]	(6.3)	−1.9[ab]	(3.7)	20[c]	(2.4)	−1.8[a]	(1.5)	23[a]	(1.5)	−0.6[a]	(0.9)
MLD	78[c]	(6.9)	−9.7[ab]	(4.1)	20[c]	(2.7)	−6.6[b]	(1.7)	12[b]	(1.6)	5.1[b]	(1.0)

Note: TA = typically achieving; LA-M = low achieving with mild retrieval deficits; LA-S = low achieving with severe retrieval deficits; MLD = mathematical learning disability. Intrusions = counting string associates of one of the addends (e.g., for 8 + 6, 9, and 7 were coded as intrusion errors). Procedural competence = (2 × frequency of min counts) + (frequency of sum counts) – (total frequency of counting errors). Group means in the same column with different superscripts differ significantly ($p < .05$).

84% and 99% of the no-choice problems were solved with retrieval; all other trials were excluded from these analyses. Mean RTs (across groups) were also consistent with use of direct retrieval; means for second, third, and fourth grade were 2662 (*SD* = 1977), 2096 (*SD* = 1635), and 1633 (*SD* = 1104) ms, respectively. Of particular interest are the percentage of retrieval errors and the percentage of these errors that were counting string intrusions (Table 3).

For the frequency of retrieval errors, the effects for group, $F(3, 219) = 25.14$, $p < .001$, and slope, $F(1, 225) = 14.28$, $p < .001$, were significant, as was the interaction, $F(3, 219) = 4.25$, $p < .01$. Follow-up contrasts confirmed the typically achieving children committed the fewest errors, followed by the LA–mild fact retrieval children ($|t|$s > 3.10, ps < .005). For the LA–severe fact retrieval children and the children with MLD, about 4 out of 5 retrieved answers were errors; these groups did not differ ($t < 1$). Although not all of the contrasts were significant, the trends for the slopes suggested the largest across-grade reductions occurred for the LA–mild fact retrieval children and the children with MLD and the smallest for the typically achieving and LA–severe fact retrieval children.

For the percentage of errors that were counting string intrusions, the effects for group, $F(1, 225) = 19.34$, slope, $F(3, 219) = 19.63$, and the interaction, $F(3, 219) = 4.77$, were significant (ps < .005). The typically achieving children had the fewest intrusion errors ($|t|$s > 2.64, ps < .01) followed by the LA–mild fact retrieval children ($|t|$s > 4.28, ps < .001), but these were common and did not differ for the LA–severe fact retrieval children and children with MLD ($p > .05$). Because of the low baseline, the slopes were shallow and did not differ for the typically achieving and LA–mild fact retrieval children ($p > .10$). In comparison, the children with MLD showed a large decrease in intrusion errors across grades ($|t|$s > 2.60, ps < .01), but the LA–severe fact retrieval children did not (ts < 1).

There are three important findings. First, the rate of retrieval errors did not differ across the LA–severe fact retrieval and MLD groups, even though the former group had many fewer retrieval errors on the choice task. The retrieval difficulties of children in the LA–severe fact retrieval group are not evident when they can resort to backup counting strategies to solve addition problems. Second, counting string intrusion errors were just as common among children in the LA–severe fact retrieval group as among the children with MLD, suggesting a common underlying deficit. Third, across second to fourth grade the overall frequency of retrieval errors and counting string intrusions declined significantly for the children with MLD, but these improvements were not found for the children in the LA–severe fact retrieval group.

Procedural competence and number tasks. To provide a contrast to the retrieval findings, we conducted the same analyses using a procedural competence variable from the addition strategy choice task, the d-prime score from the *Number Sets Test*, and the absolute error score from the number line estimation task. The procedural variable was defined as (2 × frequency of min counts) + (frequency of sum counts) – (total frequency of counting errors). High scores on this variable represent frequent and accurate use of the min procedure and low scores frequent counting errors.

For the procedural variable, the effects for group, $F(1, 186) = 16.71$, slope, $F(3, 119) = 16.04$, and the interaction, $F(3, 119) = 6.68$, were significant (ps < .001). As shown in Table 3, the children with MLD had poorly developed procedural skills in second grade ($|t|$s > 4.77, ps < .001); the three other groups did not differ (ts < 1). The interaction emerged because the procedural competence of the children with MLD improved substantially and did not differ from that of the other groups in fourth grade.

For the d-prime variable, the group, $F(1, 225) = 7.86$, slope, $F(3, 22) = 65.47$, and interaction, $F(3, 220) = 4.72$, effects were significant (ps < .01), as they were for the number line variable (Table 4), $F(1, 227) = 178.56$, $F(3, 221) = 46.95$, $F(3, 221) = 12.12$ (ps < .0001), respectively. The children with MLD scored more poorly than the other groups on

Table 4. Group Differences in Intercept and Slope for Number Sets and Number Line Tasks

	d-Prime				Absolute Number Line Error			
	Int.		Slope		Int.		Slope	
TA	0.66[a]	(0.1)	−0.1[a]	(0.1)	6[a]	(0.4)	−1.1[a]	(0.2)
LA-M	−0.25[b]	(0.1)	0.0[a]	(0.1)	8[b]	(0.4)	−1.8[b]	(0.2)
LA-S	−0.53[b]	(0.3)	0.2[a]	(0.2)	10[c]	(0.9)	−2.6[b]	(0.4)
MLD	−3.70[c]	(0.3)	0.7[b]	(0.2)	18[d]	(1.0)	−4.0[c]	(0.5)

Note: d-prime = standardized ($M = 0$, $SD = 1$), across groups and years, sensitivity to number sets; absolute number line error = absolute difference between child's placement on the number line and the correct placement; TA = typically achieving; LA-M = low achieving with mild retrieval deficits; LA-S = low achieving with severe retrieval deficits; MLD = mathematical learning disability. Group means in the same column with different superscripts differ significantly ($p < .05$).

the d-prime and number line measures in second grade ($|t|$s > 6.41, ps < .001), and children in both LA groups more poorly than the typically achieving children ($|t|$s > 4.06, ps < .01); the LA groups did not differ for d-prime ($t < 1$) but did for line number error ($t = -2.02$, $p < .05$). The interaction emerged for d-prime because of the substantial improvement of the MLD group. Their deficit remained in fourth grade, however, in comparison to the other groups, as did that of the LA groups in comparison to the typically achieving children. A similar pattern emerged for number line error, with the children in both LA groups merging in third grade and approaching that of the typically achieving children in fourth. The children with MLD improved across grades, but the gap between them and the other groups remained in fourth.

There are three core points. First, the addition fact retrieval deficits of the children in the LA–severe fact retrieval group are found with intact addition procedural competencies. Second, these children have modest deficits on the number tasks, but these are not substantially different from those of the children in LA–mild fact retrieval group, and unlike their fact retrieval errors their number competencies improve from second to fourth grade. Third, the similar fact retrieval deficits for the MLD and LA–severe fact retrieval groups and potentially more persistent deficits for the latter group are contrasted with a substantial group difference, favoring the LA–severe fact retrieval group, on the d-prime measure.

Domain-General Influences

Working memory. Group differences in working memory are shown in Table 5. Overall differences were significant for the phonological loop, $F(3, 214) = 15.94$, $p < .0001$, $d = 2.17$, visuospatial sketch pad, $F(3, 214) = 6.25$, $p < .001$, $d = 1.25$, and central executive, $F(3, 214) = 27.28$, $p < .0001$, $d = 2.67$. Follow-up honestly significant difference tests revealed the children with MLD had lower scores than all other groups on each of the working memory measures (ps < .05). The typically achieving children and the LA–severe fact retrieval children did not differ for the phonological loop or visuospatial sketch pad, but the former group had an advantage on the central executive. The typically achieving children had higher scores than the LA–mild fact retrieval children on all three measures, and the LA–mild fact retrieval and LA–severe fact retrieval groups did not differ on any of them.

Processing speed. As shown in Table 5, RAN RTs (in seconds) differed across group and grade; overall differences were significant in second, $F(3, 220) = 8.46$, $p < .0001$, $d = -1.29$, third, $F(3, 220) = 5.76$, $p < .001$, $d = -1.08$, and fourth, $F(3, 220) = 9.75$, $p < .0001$, $d = -1.27$, grades. Follow-up honestly significant difference tests revealed the children with MLD had longer RTs than the typically achieving and LA–mild fact retrieval children in each grade (ps < .05), whereas the differences between the LA–severe fact retrieval and MLD groups did not differ in any grade (ps > .05). The RTs of the LA–severe fact retrieval children did not differ from those of the typically achieving or LA–mild fact retrieval children in second or third grade but were longer in fourth grade.

The gist is that the children with MLD have pervasive working memory deficits, whereas the children in the two LA groups are average on all three core components of working memory, although the central executive scores of both of these groups are below those of their typically achieving peers. In second grade the children with MLD have very slow processing speeds, but substantially narrow the gap by fourth. The processing speeds of the two LA groups do not differ in any grade or from that of the typically achieving children, with the exception of the slower processing speed of the children in the LA–severe fact retrieval group in fourth group.

Mathematical cognition. To examine the influence of working memory, processing speed, IQ, and reading ability on the group differences on the mathematical cognition tasks, we first chose four variables from the multilevel models, specifically the children's intercept scores for percentage of intrusion errors, d-prime, number line, and procedural competence scores. We focused on intercept scores because group differences were more consistent for them than for the corresponding slopes. The scores were then standardized ($M = 0$, $SD = 1$) to put them on the same scale. Bivariate correlations revealed they were all significantly related to one another ($|r|$s = .58 to .71, ps < .0001). We then ran a repeated measures analysis of covariance, with intrusion errors, d-prime, number line, and procedural competence scores as within-subjects variables and verbal and nonverbal IQ, phonological loop, visuospatial sketch pad, central executive, and intercept values from RAN RT, word reading, and percentage of retrieval errors on the forced

Table 5. Group Differences in Working Memory and Speed of Processing

	Working Memory						Speed of Processing					
	Phonological Loop		Visuospatial Sketch Pad		Central Executive		RAN RT Second Grade		RAN RT Third Grade		RAN RT Fourth Grade	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
TA	105[a]	13	103[a]	12	106[a]	12	63[a]	11	59[a]	12	53[a]	9
LA-M	97[b]	14	98[b]	14	95[b]	13	69[a]	18	64[a]	15	56[ab]	11
LA-S	103[ab]	14	99[ab]	14	98[b]	12	70[ab]	14	66[ab]	17	62[bc]	15
MLD	80[c]	10	88[c]	12	78[c]	9	90[b]	31	80[b]	27	67[c]	13

Note: RAN RT = Rapid Automatized Naming reaction time; TA = typically achieving; LA-M = low achieving with mild retrieval deficits; LA-S = low achieving with severe retrieval deficits; MLD = mathematical learning disability. Group means in the same column with different superscripts differ significantly ($p < .05$).

retrieval task as covariates. We included the latter variable because this was used to categorize the LA children into the mild and severe fact retrieval groups, and it was therefore a necessary covariate in the analyses of the intrusion errors from the same task.

These analyses allowed us to determine if group differences on the mathematical cognition tasks remained after controlling for domain-general abilities and reading scores and if a group by mathematical cognition variable interaction emerged. If the group differences were no longer significant, then the above described differences on the mathematical tasks would be attributed to the group differences in one or several of the domain-general abilities. A significant group effect combined with a nonsignificant interaction would suggest a single mechanism underlying the group differences on the four mathematical tasks that cannot be attributed to domain-general and reading abilities, whereas a significant interaction would suggest more than one mechanism. The results revealed significant group, $F(3, 206) = 32.95$, $p < .0001$, and group by task effects, $F(9, 618) = 93.37$, $p < .0001$.

The associated least squares means, partialing the effects of the covariates, are shown in Table 6. As an example, the number of counting string intrusions of the typically achieving children was 0.75 standard deviations below the grand mean (i.e., 0), after controlling for the domain-general and reading measures. The associated t tests ($\alpha = .008$, with Bonferroni correction) revealed all of the pairwise group differences were significant ($|t|$s > 37, ps $< .001$) for the counting string intrusion variable, although the effect size for the LA–severe fact retrieval and MLD groups was small, $d = 0.10$ (i.e., 2.23–2.33). The typically achieving children had higher d-prime scores than children in the three other groups, ($|t|$s > 3.19, ps $< .001$), and children with MLD had lower scores than children in the other groups ($|t|$s > 13.90, ps $< .001$). The LA–mild fact retrieval and LA–severe fact retrieval groups did not differ ($t < 1$). For the number line and procedural competence variables, none of the comparisons of the typically achieving, LA–mild fact retrieval, and LA–severe fact retrieval groups were significant ($|t|$s < 1.83, ps $> .05$). For both tasks, the scores of the MLD children were worse than those of the children in the three other groups ($|t|$s > 5.71, ps $< .001$).

Table 6. Standardized Intercept Values From the Multilevel Models After Partialing Domain-General and Reading Abilities, and Frequency of Retrieval Errors

	Counting String Intrusions	d-Prime	Number Line Error	Procedural Competence
TA	–0.75[a]	0.43[a]	–0.20[a]	0.27[a]
LA-M	–0.01[b]	–0.04[b]	–0.08[a]	0.19[a]
LA-S	2.23[c]	–0.09[b]	–0.01[a]	–0.10[a]
MLD	2.33[d]	–2.32[c]	1.54[b]	–2.89[b]

Note: TA = typically achieving; LA-M = low achieving with mild retrieval deficits; LA-S = low achieving with severe retrieval deficits; MLD = mathematical learning disability. Values are z scores after partialing domain-general abilities. Group means in the same column with different superscripts differ significantly ($p < .05$).

The pattern for Table 6 is clear: After controlling for domain-general and reading abilities and frequency of retrieval errors on the forced retrieval task, the children with MLD have substantial deficits on all of the mathematical cognition tasks, the LA–mild fact retrieval and LA–severe fact retrieval groups have a modest deficit on the d-prime measure, and the LA–severe fact retrieval group has a substantial deficit on the counting string intrusions measure. On an individual differences level, partial correlations between these measures revealed they were still significantly related ($|pr|$s $= .25$ to $.54$, $p < .001$) but not as strongly as before partialing the domain-general, reading, and forced retrieval error variables.

Discussion

Our focus was on the most consistently found mathematical cognition deficit of children with MLD and their LA peers

(Geary, 1993; Jordan et al., 2003), specifically high error rates during the retrieval of arithmetic facts from long-term memory. The primary goal was to determine if the retrieval deficits of children with MLD are similar to or different from the deficits of children with more mild learning difficulties, the LA–severe fact retrieval group in our study, given the several distinct brain regions involved in learning and retrieving these facts (Cho et al., 2010). The inclusion of the d-prime, number line error, and procedural competence variables allowed us to assess whether the retrieval deficit co-occurs with deficits on these other quantitative tasks, as would be expected if a deficit in the ability to represent numerical magnitude cascades into deficits in mathematical competencies that are dependent on an understanding of numerical magnitude (Butterworth, 2005). Alternatively, if distinct patterns of deficit occur for different groups of children, then the implication is that there are multiple deficits or delays that can result in MLD or LA. Finally, the extensive battery of IQ, working memory, speed of processing, and reading achievement measures allowed us to assess the extent to which any mathematical cognition deficits are potentially the result of deficits in domain-general learning abilities or are specific to mathematics (also see Fuchs et al., 2010). We organize the discussion around the addition tasks and the number sets and number line tasks.

Addition Competence

The combined use of the strategy choice and forced retrieval tasks allowed us to understand retrieval deficits in the context of how children typically solve addition problems. Compared to the percentage of retrieval errors for the choice problems, children in all groups made more errors on the forced retrieval task. The across-task difference confirms predictions from Siegler's (1987, 1996) strategy choice model and empirical studies of other academic domains (Lemaire & Lecacheur, 2002), that is, that children's use of one problem-solving strategy or another results from an adaptive balance between the time needed to execute the strategy (or retrieve an answer) and the likelihood the end result will lead to the correct answer. In second grade, the ratio of retrieval errors across tasks was 12:1 and 11:1 for the typically achieving and LA–mild fact retrieval children, indicating that when these children could not correctly retrieve an answer they nearly always resorted to a backup counting strategy or use of decomposition. These children had to be highly confident of a retrieved answer, before stating it in the choice task (Siegler, 1988).

The corresponding second grade ratios for the LA–severe fact retrieval and MLD groups are 5.6:1 and 1.3:1. The latter suggests that second graders with MLD were not making adaptive choices, similar to Geary , Hoard, Byrd-Craven, and DeSoto's (2004) results for more complex addition problems. In the context of their errors rates of 20% and 31% for finger counting and verbal counting, respectively, their frequent guessing may not be as maladaptive as it seems at first blush. By fourth grade, their retrieval error ratio improved to 3.4:1, as did their procedural competence. In fourth grade, they made more adaptive choices, but still committed too many retrieval errors, given they could correctly solve these problems by counting. In second grade, the LA–severe fact retrieval children's low counting and decomposition error rates indicated that they could effectively solve many of the addition problems on which they were committing retrieval errors, suggesting a more lenient confidence criterion for stating retrieved answers. In any case, their 19:1 ratio in fourth grade indicates highly adaptive strategy choices.

The drawback of this adaptive problem solving is that the retrieval deficit of the LA–severe fact retrieval children would not be detectable using the strategy choice task, nor would it be detectable on untimed mathematics achievement tests, as children are allowed to solve the corresponding items by whatever means they choose. Their high second to fourth grade error rates on the forced retrieval task confirm Jordan et al.'s (2003) findings of a subgroup of children with specific and persistent retrieval deficits (also see Ostad, 1977; Ostad & Sorensen, 2007) that are not related to IQ and do not affect performance on at least some other mathematical tasks. Our findings take this one step further by demonstrating that the retrieval deficits of these LA–severe fact retrieval children are at least as debilitating if not more so as those of children with MLD. The finding that the retrieval errors of the children in the LA–severe fact retrieval and MLD groups are characterized by frequent counting string intrusions that are not attributable to IQ, working memory (as measured in this study), speed of processing, or reading ability suggests a similar mechanism underlying the retrieval deficits of the children in these two groups.

Although we assessed a broad range of working memory competencies, we did not assess all of them. In particular, our central executive measures largely assessed the children's ability to maintain and update information in working memory and not their inhibitory control. The latter is a subcomponent of the central executive and has been implicated as a potential contributor to MLD in the research of Passolunghi and colleagues (Passolunghi et al., 1999; Passolunghi & Siegel, 2004) as well as in Barrouillet et al.'s (1997) study of the multiplication errors of these children. This does not mean that other components of working memory do not contribute to other aspects of the mathematical learning deficits of children with MLD or some LA children, as there is evidence they do (Bull, Johnston, & Roy, 1999; Geary et al., 2007; Murphy et al., 2007; Swanson, 1993; Swanson, Jerman, & Zheng, 2008). The specific question here is whether the inhibitory component of working memory is the mechanism that is common to the retrieval deficits of children with MLD and their LA–severe

fact retrieval peers. The answer must await studies that explicitly examine the relation between this component of working memory and error rates and counting string intrusions on the forced retrieval task.

Finally, the finding that the procedural competencies of the LA–severe fact retrieval children did not differ from those of the typically achieving children supports the long-standing hypothesis of distinct mechanisms underlying procedural and retrieval deficits (Geary, 1993). The LA–severe fact retrieval children's lack of grade to grade change in the rate of forced retrieval errors and in the percentage of counting string associates suggests this deficit is not simply a developmental delay, although the pattern for the children with MLD does suggest grade to grade improvement in their ability to remember and retrieval basic facts. The grade to grade change in the children with MLD's procedural competencies is more straightforward: At least for use of counting procedures to solve simple addition problems, their early deficit represents a developmental delay of about 2 years, a gap they closed by fourth grade.

Number Sets and Number Line

The inclusion of the *Number Sets Test*, that is, the d-prime measure, and the error rate measure from the number line task allowed us to further assess the relation between fact retrieval deficits and performance in other mathematical areas. The measures do not provide a strong test of Butterworth's (2005) hypothesis that a fundamental deficit in the systems that support the exact and approximate representations of quantity are the underlying source of many persistent difficulties in learning mathematics. This is because the d-prime and number line error measures do not provide pure assessments of these systems, although they are likely dependent on them to some extent. Difficulties in subitizing and in representing approximate quantity would likely result in poor performance on the *Number Sets Test* and the number line task (Geary, Bailey, & Hoard, 2009; Siegler & Opfer, 2003), respectively. However, there are other influences on performance on these measures, such as working memory, knowledge of the formal counting system, and the ability to map Arabic numerals to associated quantities (Geary et al., 2008; Moeller, Pixner, Kaufmann, & Nuerk, 2009; Rousselle & Noël, 2007). Nevertheless, the tasks do provide at least a preliminary test of the hypothesis.

Before controlling for domain-general and reading abilities, both LA groups and the MLD group showed deficits of varying degrees on the d-prime and the number line tasks, and performance on these tasks and the counting string intrusions and procedural competence variables were highly correlated. The combination suggests there may be a single underlying core deficit, as proposed by Butterworth (2005), but other evidence suggests a more nuanced pattern of mathematical cognition deficits across groups. As noted, the procedural competence and fact retrieval deficits do not appear to be related, and once domain-general and reading abilities were controlled the deficits of the LA–mild fact retrieval and LA–severe fact retrieval groups on the number line task were no longer significant. To the extent performance on the number line task is dependent on the approximate magnitude representational system, these findings suggest that although children in the LA groups may be developmentally delayed in the maturation of this system, there does not appear to be a long-term deficit.

The fluency of these groups was still lower than that of the typically achieving children on the d-prime measure, which might result from difficulties in subitizing, in mapping Arabic numerals to associated representations of quantity, or to deficits or maturational delays in the magnitude representational systems themselves (Butterworth, 2005; Geary et al., 2009; Rousselle & Noël, 2007). Determining the source or sources of the group differences on the d-prime measures will require use of experimental measures that independently assess each of these competencies. The point for now is that there may well be individual differences in the core systems for representing and manipulating quantity that influence later mathematical learning, as suggested by Butterworth, but deficits in any such core systems do not appear to be a sufficient explanation for the pattern and developmental trajectory of the mathematical learning difficulties exhibited by LA–severe fact retrieval children and children with MLD.

Summary

The prospective design of the study and administration of the strategy choice and forced addition fact retrieval tasks as well as other mathematical tasks and IQ, working memory, processing speed, and reading measures allowed us to confirm Jordan et al.'s (2003) finding of a subgroup of LA children with persistent arithmetic fact retrieval deficits. We extended these findings by demonstrating these deficits are at least as severe as those of children with MLD, that is, children with broader deficits in mathematical cognition and general learning abilities. The deficits of these LA–severe fact retrieval children are characterized by frequent retrieval of counting string associates while attempting to remember addition facts, suggesting poor inhibition of irrelevant information into working memory during the retrieval process (Passolunghi & Siegel, 2004), although this hypothesis remains to be directly tested. In any event, the deficits of these children are not detectable by untimed standard mathematics achievement tests (they may be detectable with standardized fluency measures) and occur despite average verbal and nonverbal intelligence and reading ability and average performance on many standard working memory tasks.

Acknowledgments

We thank Linda Coutts, Lara Nugent, Jennifer Byrd-Craven, Rachel Christensen, Caitlin Cole, Mike Coutts, Patricia Hoard, Rehab Mojid, Cy Nadler, Robyn Tenney, Jonathan Thacker, Leah Thomas, Jasmine Tilghman, Chatty Numtee, Erin Twellman, Katie Waller, and Jessica Wallace for help on various aspects of the project.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The authors received the following financial support for the research and/or authorship of this article:

Geary acknowledges support from Grant R01 HD38283 from the National Institute of Child Health and Human Development (NICHD) and Grant R37 HD045914 cofunded by NICHD and the Institute of Educational Sciences.

Note

1. The overall effect size (*d*) was the difference between the maximum and minimum group mean divided by the overall standard deviation (Cohen, 1988).

References

Barrouillet, P., Fayol, M., & Lathuliére, E. (1997). Selecting between competitors in multiplication tasks: An explanation of the errors produced by adolescents with learning disabilities. *International Journal of Behavioral Development*, *21*, 253–275.

Berch, D. B., & Mazzocco, M. M. M. (Eds.). (2007). *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities*. Baltimore, MD: Brookes.

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

Bull, R., Johnston, R. S., & Roy, J. A. (1999). Exploring the roles of the visual-spatial sketch pad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology*, *15*, 421–442.

Butterworth, B. (2005). Developmental dyscalculia. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 455–467). New York, NY: Psychology Press.

Butterworth, B., & Reigosa, V. (2007). Information processing deficits in dyscalculia. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 65–81). Baltimore, MD: Brookes.

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*, *9*, 613–625.

Cho, S., Ryali, S., Geary, D. C., & Menon, V. (2010). *How do you solve 7 + 8? Decoding children's brain activity patterns during counting versus retrieval*. Manuscript submitted for publication.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Denckla, M. B., & Rudel, R. (1976). Rapid Automatized Naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*, 471–479.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

Every Child a Chance Trust. (2009). *The long-term costs of numeracy difficulties*. Retrieved from http://www.everychildachancetrust.org/counts/index.cfm

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

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., & Hamlett, C. L. (2010). The contributions of numerosity and domain-general abilities for school readiness. *Child Development*, *81*, 1520–1533.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology*, *49*, 363–383.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*, 345–362.

Geary, D. C. (2010). Missouri Longitudinal Study of Mathematical Development and Disability. *British Journal of Educational Psychology Monograph Series II*, *7*, 31–49.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The Number Sets Test. *Journal of Psychoeducational Assessment*, *27*, 265–279.

Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., & Nugent, L. (2009). First-grade predictors of mathematical learning disability: A latent class trajectory analysis. *Cognitive Development*, *34*, 411–429.

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology*, *77*, 236–263.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, C. M. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, *88*, 121-151.

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

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

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, *79*, 329–343.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*, *85*, 103–119.

Jordan, N. C., & Montani, T. O. (1997). Cognitive arithmetic and problem solving: A comparison of children with specific and general mathematics difficulties. *Journal of Learning Disabilities*, *30*, 624–634.

Kadosh, R. C., Kadosh, K. C., Schuhmann, T., Kaas, A., Goebel, R., Henik, A., & Sack, A. T. (2007). Virtual dyscalculia induced by parietal-lobe TMS impairs automatic magnitude processing. *Current Biology*, *17*, 689–693.

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition*, *2*, 1–23.

Lemaire, P., & Lecacheur, M. (2002). Applying the choice/no-choice methodology: The case of children's strategy use in spelling. *Developmental Science*, *5*, 42–47.

MacMillan, N. A. (2002). Signal detection theory. In J. Wixted & H. Pashler (Eds.), H. *Stevens' handbook of experimental psychology*, *3rd edition*, *vol. 4: Methodology in experimental psychology* (pp. 43–90). New York, NY: John Wiley.

Mazzocco, M. M., & Myers, G. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia*, *53*, 218–253.

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology*, *103*, 503–515.

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities*, *40*, 458–478.

Ostad, S. A. (1997). Developmental differences in addition strategies: A comparison of mathematically disabled and mathematically normal children. *British Journal of Educational Psychology*, *67*, 345–357.

Ostad, S. A., & Sorensen, P. M. (2007). Private speech and strategy-use patterns: Bidirectional comparisons of children with and without mathematical difficulties in a developmental perspective. *Journal of Learning Disabilities*, *40*, 2–14.

Passolunghi, M. C., Cornoldi, C., & De Liberto, S. (1999). Working memory and intrusions of irrelevant information in a group of specific poor problem solvers. *Memory & Cognition*, *27*, 779–790.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, *88*, 348–367.

Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children (WMTB-C) manual*. London, UK: Psychological Corporation.

Raven, J. C., Court, J. H., & Raven, J. (1993). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London, UK: H. K. Lewis.

Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex*, *15*, 1779–1790.

Rivera-Batiz, F. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources*, *27*, 313–328.

Rousselle, L., & Noël, M.-P. (2007). Basic numerical skills in children with mathematical learning disabilities: A comparison of symbolic vs. non-symbolic number magnitude processing. *Cognition*, *102*, 361–395.

Russell, R. L., & Ginsburg, H. P. (1984). Cognitive analysis of children's mathematical difficulties. *Cognition and Instruction*, *1*, 217–244.

SAS Institute. (2004). Statistical analysis system 7.5. Cary, NC: Author.

Scott Foresman. (1999). *Investigations in number*, *data*, *and space*. Glenview, IL: Author.

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

Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development*, *59*, 833–851.

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

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

Siegler, R. S., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243.

Siegler, R. S., & Shrager, J. (1984). Strategy choice in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Lawrence Erlbaum.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, *27*, 279–306.

Starkey, P., & Cooper, R. G., Jr. (1980, November 28). Perception of numbers by human infants. *Science*, *210*, 1033–1035.

Swanson, H. L. (1993). Working memory in learning disability subgroups. *Journal of Experimental Child Psychology*, *56*, 87–114.

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, *100*, 343–379.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are

important. *Journal of Experimental Child Psychology*, *79*, 294–321.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation, Harcourt Brace.

Wechsler, D. (2001). *Wechsler Individual Achievement Test–II–Abbreviated*. San Antonio, TX: Psychological Corporation, Harcourt Brace.

About the Authors

David C. Geary, PhD, is a Curators' Professor and Thomas Jefferson Professor at the University of Missouri. He was a member of the President's National Mathematics Advisory Panel and, among other things, is interested in individual differences in children's mathematical cognition and learning.

Mary K. Hoard, PhD, is a Senior Research Specialist at the University of Missouri. Her current interests include mathematical development and cognition and chronobiology.

Drew H. Bailey, MA, is a graduate student at the University of Missouri. He is interested in, among other things, mathematical cognition and development.