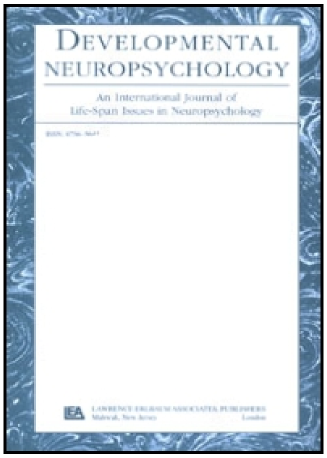

Developmental Neuropsychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653638

Standardized Assessment of Strategy Use and Working Memory in Early Mental Arithmetic Performance

Sarah S. Wu [a]; Meghan L. Meyer [a]; Uta Maeda [a]; Valorie Salimpoor [b]; Sylvia Tomiyama [c]; David C. Geary [d]; Vinod Menon [e]
[a] Department of Psychiatry and Behavioral Sciences, Stanford University,
[b] Department of Psychiatry, McGill University, Montreal, Canada
[c] Department of Psychiatry, Stanford University,
[d] Department of Psychological Sciences, University of Missouri,
[e] Department of Psychiatry and Behavioral Sciences, Symbolic Systems Program, Program in Neuroscience, Stanford University,

Online Publication Date: 01 May 2008

To cite this Article: Wu, Sarah S., Meyer, Meghan L., Maeda, Uta, Salimpoor, Valorie, Tomiyama, Sylvia, Geary, David C. and Menon, Vinod (2008) 'Standardized Assessment of Strategy Use and Working Memory in Early Mental Arithmetic Performance', Developmental Neuropsychology, 33:3, 365 — 393

To link to this article: DOI: 10.1080/87565640801982445
URL: http://dx.doi.org/10.1080/87565640801982445

DEVELOPMENTAL NEUROPSYCHOLOGY, *33*(3), 365–393

ISSN: 8756-5641 print / 1532-6942 online
DOI: 10.1080/87565640801982445

Standardized Assessment of Strategy Use and Working Memory in Early Mental Arithmetic Performance

Sarah S. Wu, Meghan L. Meyer, and Uta Maeda
Department of Psychiatry and Behavioral Science
Stanford University

Valorie Salimpoor
Department of Psychology
McGill University, Montreal, Canada

Sylvia Tomiyama
Department of Psychology
Stanford University

David C. Geary
Department of Psychological Sciences
University of Missouri

Vinod Menon
Department of Psychiatry and Behavioral Science
Symbolic Systems Program, Program in Neuroscience
Stanford University

Although children's use of a variety of strategies to solve arithmetic problems has been well documented, there is no agreed on standardized and validated method for assessing this mix. We examined the convergent validity of typically achieving (TA,

Correspondence should be addressed to Vinod Menon, Symbolic Systems Program, Program in Neuroscience and Department of Psychiatry & Behavioral Sciences, 780 Welch Rd., Room 201, Stanford University School of Medicine, Stanford, CA 94305-5778. E-mail: menon@stanford.edu

$N = 39$) and low achieving (LA, $N = 20$) second and third grade children's strategy choices in simple addition using three different methods: child self-report, observer-report, and response time (RT). The high concordance between child and observer reports (Kappa = .948) in both groups suggests that the participants were aware of, and could accurately report, the strategies they used. The Receiver-Operator Characteristic (ROC) analysis showed that RT accurately differentiated between retrieval and counting (AUC = 82%). The specificity and sensitivity of the ROC profiles were significantly greater for the TA group than for LA group, even though the groups did not differ in the overall strategy mix. Our findings suggest that ROC analysis is more sensitive to group differences in the mechanisms governing strategy choice than observation or child report. Children's use of retrieval strategies as well as accuracy during both retrieval and counting trials were all related to the central executive, but not the phonological and visuospatial sketchpad, component of working memory. We discuss the implication of these findings for early mathematical learning.

It is well established that children's problem solving in many cognitive and academic domains is characterized by use of a mix of strategies and that development involves gradual change in this mix. The change is reflected in declining use of the least efficient—in terms of time, accuracy, and working memory demands—strategies and increasing use of the most efficient strategies (Siegler & Shrager, 1984). One of the areas in which this pattern has been extensively documented is in the mix of strategies children use to solve arithmetic problems. These studies have shown that even young children are capable of using a variety of strategies to solve simple addition and subtraction problems, and that the strategy mix changes with development and schooling (Ashcraft, 1992; Baroody, 1987; Carpenter & Moser, 1984; Cooney, Swanson, & Ladd, 1988; Siegler & Robinson, 1982; Siegler & Shrager, 1984). The developmental pattern is most evident in typically achieving (TA) children who progress from the use of immature counting strategies to the more mature memory-based processes during the elementary school years (Ashcraft & Fierman, 1982; Geary, Widaman, Little, & Cormier, 1987; Kaye, Post, Hall, & Dineen, 1986). Children with a mathematical learning disability (MLD) and, to a lesser extent, children who score lower on mathematics achievement tests than would be expected on their intelligence and reading scores (i.e., low achieving children[LA]) use counting strategies for more years and show evidence of difficulties in use of memory-based processes (Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Geary & Brown, 1991).

Researchers have employed a variety of methods to study the mix of strategies children use to solve arithmetic problems, but to date there is no agreed-on standardized measures for assessing these strategy choices. In fact, there have been concerns regarding the validity of the commonly used method of subject report and whether these reports are consistent with RT patterns (Ashcraft, 1992). We address this issue in the current study with the assessment of the convergent validity of the three methods that have been used for assessing strategy choices in arithmetic for groups of typically achieving and low achieving children; specifically, experimenter observation, child self-report, and RT. We also present the first Receiver-Operator Characteristic (ROC) analysis of RT measures of strategy choice and relate these parameters and those derived from the observational and self-report methods to individual and group differences in working memory.

ADDITION STRATEGY CHOICES

During the early elementary school years, the mix of strategies children use to solve simple addition problem is initially dominated by finger and verbal counting, although most children can use direct retrieval to solve a few problems, such as retrieving "three" for 1 + 2 (Ashcraft, 1982; Fuson, 1982; Groen & Parkman, 1972; Siegler & Shrager, 1984). With schooling, the use of counting gradually declines and is replaced by direct retrieval of answers from long-term memory, at least for simple problems.

In comparison to typically achieving children, children with MLD and their low achieving peers use counting as a problem solving strategy across more grades and commit more counting errors (Geary, 1993; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Hanich, Jordan, Kaplan, & Dick, 2001; Jordan & Montani, 1997; Ostad, 1997; Russell & Ginsberg, 1984). The most consistent finding is that children with MLD and, to a lesser extent, low achieving children show a deficit in the ability to use retrieval-based processes (Barrouillet, Fayol, & Lathuliére, 1997; Geary,1990; Geary, Hamson, & Hoard, 2000; Jordan, Hanich, & Kaplan, 2003). It is not that these children never correctly retrieve answers, but that they show a persistent difference in the frequency with which they correctly retrieve basic facts, and sometimes in the pattern of retrieval errors.

STANDARDIZING STRATEGY USE ASSESSMENTS

To date, three methods have been used to evaluate strategy choices. The simplest of these is the self-report, which involves directly asking children which strategy they used to solve each problem (Carpenter & Moser, 1984; Houlihan & Ginsburg, 1981). The assumption is that children are capable of providing an accurate description of their strategies, if asked immediately after solving each problem (Siegler, 1987; Svenson, 1985). Observer-reports require the experimenter to watch a child solve each problem and note overt signs of counting, such as finger usage, lip movement, or audible counting. However, this type of classification becomes increasingly more difficult with age, because as children progress to implicit counting and to retrieval, overt signs of strategy use begin to disappear.

A third method is chronometry, that is, the measurement of RT for solving each problem (Groen & Parkman, 1972), and use of regression equations to infer underlying strategies and processes. For instance, if problems are solved by means of counting, then RT should be linearly related to the magnitude of the counted addend and the corresponding raw regression coefficient should be consistent with independent estimates of speed of implicit counting (Geary & Widaman, 1987; Groen & Parkman, 1972; Svenson, 1985; Widaman, Little, Geary & Cormier, 1992). For retrieval, the best predictor of RTs should be a variable that models the underlying structure of the long-term memory organization of addition facts (Ashcraft & Battaglia, 1978). The major advantages of the chronometric model are that the RTs can be reliably assessed, they are not dependent on subject report, and do not lose their sensitivity as observation of strategy choices becomes difficult (e.g., for retrieval). The primary disadvantage is that it is difficult to assess within-subject variation in strategy usage.

In this study, we used ROC methods, a signal detection method that determines how well a predictive equation correctly classifies a data set into two groups (Centor, 1991). In the domain of mathematical cognition, ROC analyses have been used to predict performance outcomes in early mathematics learning. For example, Mazzocco and Thompson (2005) used ROC curves to determine whether performance measures gathered in kindergarten could predict whether participants would eventually develop MLD in the second and third grade. Here we use ROC methods to assess whether RTs generated when second- and third-grade children solve additional problems can be used to accurately classify strategy choices, more specifically, to distinguish between retrieval and counting. We also determined the RT, or cutoff criterion, that maximizes the sensitivity and specificity of the classification. These analyses may be used to help develop a confirmatory and quantitative method of using RTs for determining strategy use.

Thus, the first set of aims of the current study are to: (1) test the level of convergence among experimenter reports and participant self-reports of children's strategy choice, (2) validate the aforementioned qualitative methods of strategy assessment with chronometric data, (3) investigate a predictive model of strategy use based on RTs, (4) determine a RT cutoff threshold that maximizes the differentiability between the two strategies of counting and retrieval, and (5) compare strategy use in typically and low achieving children using all of these methods.

STRATEGY USE AND WORKING MEMORY

The second aspect of our study concerns the working memory contributions to individual and group differences in children's strategy choices. On the basis of

Baddeley and Hitch's (1974) model, the three key components of working memory are the central executive, phonological loop, and visuospatial sketchpad. The central executive functions as a supervisory and control system, orchestrating the activity and flow of information represented in the phonological and visuospatial subsystems. The central executive is also engaged in other aspects of executive function more generally (Baddeley, 1996) including retrieval of information from long-term memory (Baddeley, 1996; Baddeley, Emslie, Kolodny, & Duncan, 1998). The phonological loop and visuospatial sketchpad act as short-term storage systems for verbal and visuospatial information, respectively (Baddeley, Thomson, & Buchanan, 1975; Gathercole & Adams, 1994; Della Sala, Gray, Baddeley, Allamano, & Wilson,1999; Logie, 1986; Wilson & Swanson, 2001).

With respect to how children solve addition problems, the central executive can influence strategy choice in both task-specific and task-general ways (Menon, Meyer, & Wu, in press). In general, this component of working memory can play an important role in making decisions about appropriate strategy use and then allocating attentional resources to implement the selected strategy (Shrager & Siegler, 1998). The role of the central executive may be particularly critical in second- and third-grade children, when these tasks are not well automated and effortful processing is required. If an answer cannot be retrieved from long-tem memory, the central executive must engage several sequential processing stages to perform the task. These include encoding, maintenance, and manipulation of numbers in either the visuospatial and phonological buffers (Geary & Brown, 1991; Hitch, 1978; Logie, Gilhooly, & Wynn, 1994; Wilson & Swanson, 2001).

Previous research has established that individual and group differences in mathematics achievement are related to working memory capacity (Bull, Johnston, & Roy, 1999; Geary, Hoard, Byrd-Craven, & DeSoto, 2004; McLean & Hitch, 1999; Swanson, 1993, 2006; Swanson & Sachse-Lee, 2001); Barrouillet and Lepine (2005) found that third- and fourth-grade children with high working memory capacity, as determined by dot-counting and reading letter span tasks, retrieved answers to simple addition questions (e.g., 3 + 4 = 7) significantly more frequently and more quickly than did children with low working memory capacities. Similarly, Bull, Johnston, and Roy (1999) found a significant, negative correlation between frequency of arithmetic fact retrieval and perseveration on the Wisconsin Card Sorting Test, a measure of the central executive component of working memory in third-grade children. Geary et al. (2004) found that the poor central executive capacity (as measured by a counting span task) was associated with more frequent use of counting to solve addition problems and more counting errors in first graders. This pattern was not found in the third graders, suggesting a developmental shift in the importance of the central executive for solving simple addition problems sometime between first and third grade. The age at which this shift predominantly occurs is not known as few studies have examined strategy use and the role of working memory in second-grade children.

Despite these advances, the components of working memory that contribute to this relation are not as well understood. For example, Barrouillet and Lepine (2005) used a measure for the central executive (i.e., the Dot-Counting task), but did not include standardized measures for the phonological loop and the visuospatial sketchpad. Bull et al. (1999) assessed both the central executive and visuospatial sketchpad, but these measures were not standardized. In addition, they did not examine the role of the phonological component of working memory, although this component has been implicated as a strong predictor of first-grade children's arithmetic performance (Rasmussen & Bisanz, 2005). However, Geary et al. (2007) recently demonstrated that, in mathematics achievement and in an array of mathematical cognition tasks, including an assessment of strategy choices in addition, individual and group differences were consistently related to the individual differences in the capacity of the central executive. The phonological loop and visuospatial sketchpad, in contrast, contributed to more circumscribed aspects of mathematical performance. We provide a follow up to this study by examining the relation between each of the three core components of working memory and addition strategy choices assessed by the three different methods described earlier. We also examine the role of each working memory component in the accurate and efficient performance on the addition trials in the TA and LA groups.

METHOD

Participants

Participants were 59 second ($N = 26$) or third ($N = 33$) graders recruited from the San Francisco Bay Area. Participants included 31 boys and 28 girls with a mean age of 8.05 years ($SD = 0.66$) at Time 1 testing. Advertisements in school and local newspapers, community and electronic bulletin boards, community organizations, and other public locations were the principal means of recruitment. Parents who wished to enroll their children in the study contacted the lab for initial screening and further information. Written consent was obtained from all participants and parents prior to any participation.

Only children with IQ scores between 75 and 125 were included. Our classification of TA, LA, and MLD groups was based on the approach outlined by Mazzocco in the introduction to this special issue. Children were classified as typically achieving ($N = 31$) if either their standardized Numerical Operations or their Math Composite score from the *Weschler Individual Achievement Test—II* (*WIAT-II*; Weschler, 2005) score was greater than 90 (corresponding to the 25th percentile). Children were classified as low achieving ($N = 28$) if either score was less than 90 but greater than 82 (corresponding to the 12th percentile). According to the three group criteria, children who scored below the 12th percentile are to be classified as MLD. However, only one child in our study met the recommended stricter criteria for MLD, so no MLD/LA distinction is made here and the group as a whole is referred to as "low achieving." However, prior to the recommendation to use the three groups most studies would have classified these children as "MLD." An independent sample *t*-test indicated no differences in age or grades between the typically and low achieving participants.

Measures

Intelligence was assessed using the *Weschler Abbreviated Scales of Intelligence* (*WASI*; Weschler, 1999). The Numerical Operations, Math Reasoning, and Word Reading subtests of the WIAT-II were used to assess achievement.

Working memory was assessed using four subtests of the *Working Memory Test Battery for Children* (*WMTB-C*; Pickering & Gathercole, 2001): Counting Recall, Backward Digit Recall, Digit Recall, and Block Recall. Counting Recall and Backward Digit Recall assess the central executive component of working memory, whereas Digit Recall and Block Recall assess phonological loop and visuospatial sketchpad components, respectively. Counting Recall requires the child to count a set of 4, 5, 6, or 7 dots on a card, and then to recall the number of counted dots at the end of a series of cards. Backward Digit Recall is a standard format Backward Digit Span. Digit Recall is a span task in which the child is asked to repeat number words spoken by the experimenter in the same order as presented by the experimenter. 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 only be seen from the experimenter's perspective. The experimenter taps a block (or series of blocks), and the child's task is to repeat the tapping in the same order as presented by the experimenter.

Strategy assessment. The stimuli included 24 addition problems in the format of "a + b = ?" composed of random pairs of integers 2 to 19, with sums ranging from 6 to 25. Tie problems (e.g., 2 + 2, 5 + 5) and addends of 0 and 1 were excluded because answers are typically retrieved from long-term memory and thus do not show strategy variability even for young children (Groen & Parkman, 1972; Svenson, 1985). No repetition of either addend was allowed across consecutive problems. Four practice problems were provided at the beginning of each session in order to familiarize the participants with the task.

Procedure

Strategy assessment. Following 4 practice problems, participants were asked to solve the same 24 addition problems, presented one at a time on the computer screen. The child was asked to solve each problem (without the use of paper and pencil) as quickly as possible without making too many mistakes and to immediately state the answer out loud. It was emphasized that the child could use whatever strategy was easiest to produce an answer; that is, they were told that they could count using their fingers, count verbally, or "just remember the answer." For each problem, the experimenter took detailed notes of overt signs of counting, such as finger usage, lip movement, or audible counting. A timer was started at the initial display of each problem, and the experimenter measured the child's RT by pressing a key on the keyboard as soon as a verbal response was given; all sessions were audio-recorded to check for RT precision. Immediately after each response, children were asked what strategy they used to solve the problem. Their reported strategies were classified as one of the four basic types: (a) counting fingers, (b) fingers (with no visible or audible verbal counting), (c) verbal counting, and (d) retrieval (Geary & Burlingham-Dupree, 1989; Geary, Hoard, Nugent, Byrd-Craven, 2007; Siegler & Shrager, 1984; Siegler, 1986). The experimenter's observations and the participant's descriptions were compared to determine the degree of agreement.

All strategies were reclassified as either "counting" (counting on fingers, finger usage without verbal counting, and verbal counting) or "retrieval." All other strategies, such as decomposition, guessing, or counting by numbers (e.g., counting by 2) comprised approximately 7.14% of the total number of trials, and were excluded from further analysis. Trials that the children reported as having been solved by guessing were removed in order to retain only trials actively solved by retrieval. If it was unclear whether the child actively attempted to retrieve the answer or guessed, the trial was removed. Isolated trials (33 total) in which the RT was flawed due to the experimenter pressing the key too early or too late were removed from the analysis, as were trials with an RT > 3.5 *SD*s above or below the mean. Participants who counted on at least 60% of the trials were classified as "counters," and likewise, participants who retrieved on at least 60% of the trials were considered "retrievers." Those who used a roughly equal proportion of counting and retrieval were classified as "neither."

ROC. To investigate the appropriateness of using RT to determine strategy, we used Receiver Operating Characteristic (ROC), a method of signal detection. ROC assesses how well a variable of interest classifies a data set into two groups (Centor, 1991; Katz & Foxman, 1993). For any particular threshold, all values above are classified into one group, whereas all values below are classified into a second group. Each threshold yields an associated proportion of cases that are correctly classified in a group (i.e., "sensitivity," or percentage true positives), and the proportion that are incorrectly classified into the same group (i.e., 1 minus the "specificity," or percentage false positives). The threshold is then varied to yield an ROC curve (Center, 1991). The ROC curve is an X-Y plot of all the sensitivity and specificity pairs in a unit square (i.e. maximum X and Y values are both 1). The

ROC curve can be used to determine the threshold point at which the sensitivity and specificity is optimal; thus maximizing the percentage of true positives while minimizing the percentage of false positives. ROC curves also provide a measure called the Area under the Curve (AUC), or, the likelihood that, given a true positive and a true negative, the ROC would correctly classify each as such (Hanley & McNeil, 1982; Lusted, 1971). An AUC of .5, for example, would indicate that the ROC was unreliable, assigning a correct classification in only 50% of the cases. An AUC of 1, however, indicates that the ROC was perfectly accurate in classification. In our study we used response times to differentiate between counting and retrieval trials. True positives here refer to the percentage of counting trials that were accurately identified as counting trials, and false positives refer to the percentage of retrieval trials that were inaccurately identified as counting trials.

RESULTS

The results are presented in three sections. In the first section, we present results from the analysis of the correspondence between self- and observer-reports. In the second section, we present results from the ROC analyses, first with the entire group of participants, and then between-group comparisons of ROC curves. In the third section, we also present results detailing behavioral measures of performance stemming from strategy use. Finally, we discuss the relationship between strategy choice and the three components of working memory.

As expected, the typically achieving (N = 39) participants had significantly greater WIAT-II Numerical Operations and Math Composite scores than the low achieving (N = 20) participants (Table 1). Both groups had average and statistically similar WIAT word reading and reading comprehension scores that were well above the reading disability cutoff score of 91.

Correspondence Between Self- and Observer-Reports

Our analysis revealed a Kappa's α of .948 ($p < .001$) between the children's self-report and experimenter's observations of strategy use. The most common disagreement occurred when the child reported using retrieval but the experimenter observed subtle signs of counting, such as fingers fluttering or lips moving. The Kappa's α was .944 and .954 (both $p < .001$) for the TA and LA groups, respectively.

ROC Profiles

Across all children, the ROC analysis yielded an AUC of 0.820 (SE = 0.012, asymptotic $p < .001$; Figure 1), suggesting that RT could successfully differentiate

TABLE 1
Neuropsychological Assessment Comparisons of Typically Achieving and Low Achieving Children

	Group					
	Typically Achieving		*Low Achieving*		*t-Test*	
Measure	*M*	*SD*	*M*	*SD*	*t-score*	*P*
WASI						
Verbal	108.62	13.22	113.36	14.60	–1.28	*ns*
Performance	106.21	14.01	105.32	12.12	0.25	*ns*
Full	108.05	11.61	110.45	12.03	–0.766	*ns*
WIAT						
Word Reading	109.20	10.47	106.14	13.65	0.98	*ns*
Reading Comprehension	105.64	10.21	104.73	10.22	0.74	*ns*
Numerical Operations	111.74	10.00	92.14	12.52	6.71	<.000
Math Reasoning	110.13	10.89	103.27	11.51	2.31	<.000
Math Composite	112.49	10.37	96.73	11.20	5.54	<.000
WMBTC						
Digit Recall	105.92	16.63	107.32	18.27	–0.30	*ns*
Backwards Recall	98.00	12.54	90.59	11.96	2.24	0.029
Counting Recall	98.34	16.06	75.14	17.97	2.72	0.009
Block Recall	95.82	15.96	96.86	18.40	–0.23	*ns*
Strategy Assessment						
Percentage Retrieved[a]	0.52	0.29	0.38	0.20	1.97	0.054
Average Response Time[b]	4476.59	1511.78	5481.69	1559.78	–2.39	0.02

Note: Group consisted of 39 Typically Achieving and 20 Low Achieving children. [a]Percentage retrieved out of total trials. [b]Measurement in milliseconds.

between the trials in which participants used counting and trials in which they used retrieval. The ROC data yielded a maximal cutoff of 3662.5 ms, with a sensitivity of 78.7% and a specificity of 72.1%. In other words, trials with RTs of less than 3662.5 ms most likely reflect the use of retrieval, whereas trials with RTs greater than 3662.5 ms reflect the use of counting. In using this cutoff threshold, 78.7% of the counting trials will be correctly classified as being solved by counting, whereas 27.9% (i.e., 100%–72.1%) of the retrieval trials will have been incorrectly categorized as having been solved by counting.

Comparison of second and third graders. For the second graders, the ROC analysis yielded an AUC of .801 ($SE = 0.019$), with a cutoff RT of 4047 ms, 77.4% sensitivity, and 72.5% specificity. The third grade ROC data yielded an AUC of .831 ($SE = 0.015$), a cutoff RT of 3426 ms, 79.3% sensitivity, and 73.8% specificity. A z-score analysis (Hanley & McNeil, 1982) indicated no significant differences between the two curves ($z = 1.198$, $p = .231$).

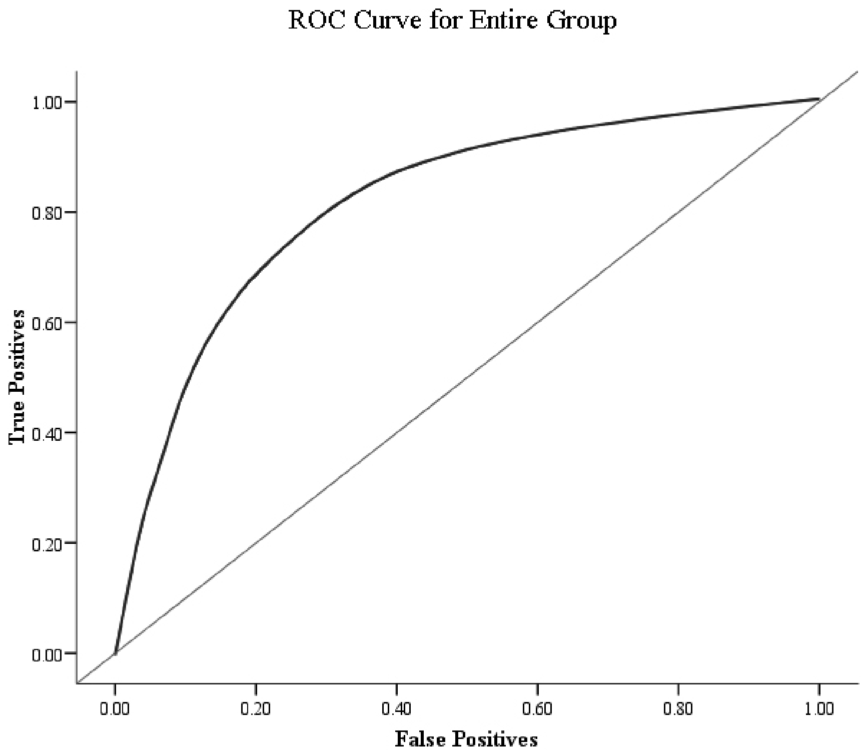

FIGURE 1 ROC Curve for the entire group. The AUC of 0.820 suggested that RT successfully differentiated between counting and retrieval methods.

Comparison of typically and low achieving groups. The ROC analysis yielded an AUC of .839 for the TA group, and an AUC of .756 for the LA group. For the TA group, the maximal cutoff RT was 3603 ms, which yielded a sensitivity of 77.6% and a specificity of 77.1%. For the LA group, the maximal cutoff RT was 4340 ms, with a sensitivity of 74.1% and a specificity of 69.5%. A z-score analysis indicated that the AUC was significantly greater for the TA group than the LA group ($z = 2.897$, $p = .004$; Figure 2).

Strategy Use and Behavioral Performance

Within the TA group, 10 children (25.6%) were retrievers, 19 were counters (48.7%), and 10 used the two strategies equally (25.6%). Within the LA group, 4 children (20%) were retrievers, 9 (45%) were counters, and 7 (35%) used the two strategies equally. A chi-square test indicated no significant differences within the

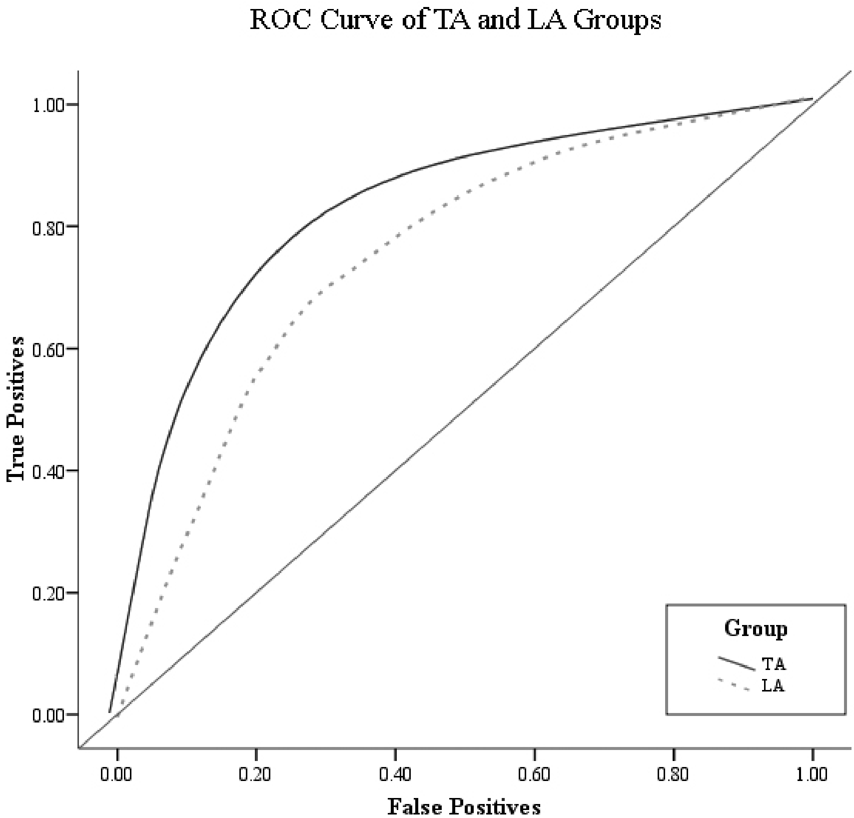

FIGURE 2 ROC Curve for the typically achieving (TA) and low achieving (LA) groups. Area under the curve (AUC) for the TA group was significantly greater than the LA group, suggesting that the measurement of strategy assessment was more accurate in the TA group.

TA and LA groups in the distributions of retrievers, counters, or children who used both equally ($\chi^2[2] = 0.618$, $p = .734$). The findings, however, suggested that typically achieving children tended to rely more on retrieval (typically achieving, $M =$ 52%, $SD = 29\%$; low achieving, M = 38%, $SD = 20\%$), whereas low achieving children relied more on counting $F(1, 57) = 3.881$, $p = .054$.

Strategy use and RT. Across groups, average RTs were longer for counting trials than for retrieval trials, $F(1, 53) = 104.776$, $p < .001$, and this was also the case for each group; $F(1, 33) = 63.813$, $p < .001$, and $F(1, 19) = 40.487$, $p < .001$, respectively for TA and LA groups (Figure 3). Overall, children in the typically achieving group had faster RTs than the low achieving participants (F[1, 57] =

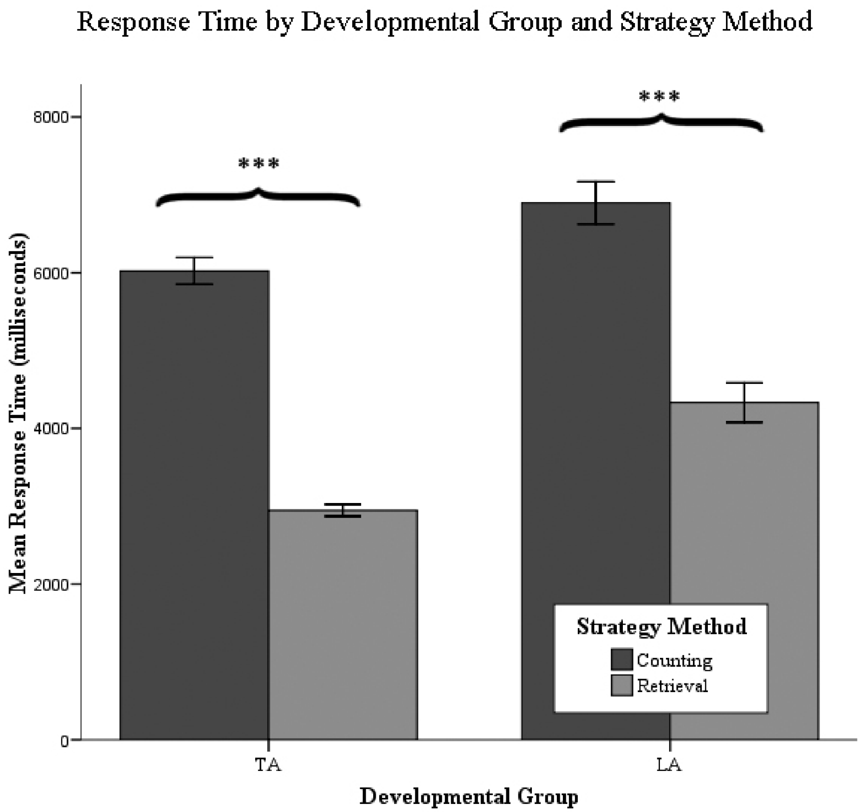

FIGURE 3 Reaction times were significantly greater during counting trials than retrieval trials for both typically achieving (TA) and low achieving (LA) participants. Bars indicate ±1 Standard Error. ***indicates $p < .001$.

5.721, $p < .020$; Table 1). An independent samples t-test indicated that typically and low achieving children did not differ in their counting RTs, $t(52) = -.912$, $p = .336$. However, typically achieving children were faster than the low achieving children in retrieving, $t(57) = -3.523$, $p = .001$.

Although the within-group regressions were non-significant, the data appear to indicate between-group differences in the strategy choice and RT profiles (Figure 4; typically achieving, standardized $\beta = -.220$, $p =$ ns, $R^2 = .047$; low achieving, standardized $\beta = .296$, $p =$ ns, $R^2 = .087$). A difference of slope test indicated that the slopes of RT regressed on percentage retrieved were significantly different (standardized $\beta = .346$, $t[1] = 2.706$, $p = .007$) for the two groups. This suggests that the RT profiles relative to strategy use for the two groups were significantly different. Typically achieving children become faster as they retrieve more, where-

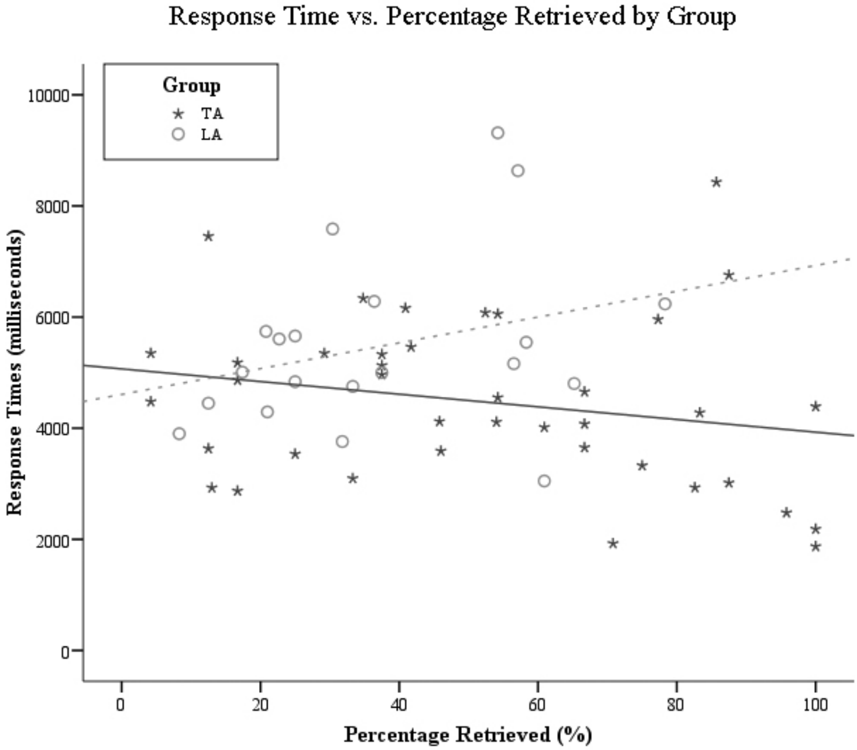

FIGURE 4 Reaction time profiles differed between the typically achieving (TA) and low achieving (LA) groups. The TA group became faster with greater reliance on retrieval, whereas the LA children group showed an opposite profile. The data suggest that the LA group was less proficient in retrieval than the TA group.

as low achieving children exhibit a significantly different profile, becoming slower as they depend more heavily on retrieval.

Strategy use and accuracy. Accuracy for counting and retrieval trials did not differ, $F(1, 53) = 3.431$, $p = .070$. The typically achieving children were more accurate overall (96.14% correct) than their low achieving peers (89.97% correct), $F(1, 57) = 11.996$, $p = .001$. An independent samples *t*-test indicated that typically achieving children were more accurate than low achieving children in both counting, $t(52) = 2.369$, $p = .022$, and retrieving, $t(57) = 2.492$, $p = .016$. Within each group, the children were equally accurate for counting and retrieval trials; typically achieving, $F(1, 33) = 2.073$, $p = .159$, low achieving, $F(1, 19) = 1.428$, $p = .247$; Figure 5.

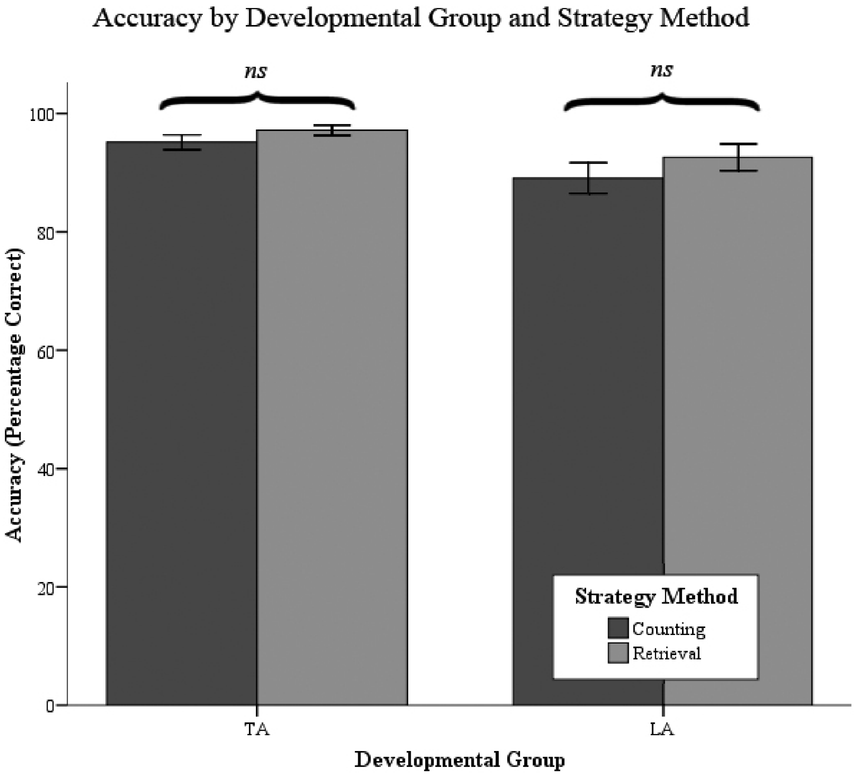

FIGURE 5 Both the TA and LA groups were equally accurate in counting and retrieval. Bars indicate ±1 Standard Error.

RT and sum size. Across groups, average RTs increased linearly with sum size, standardized $\beta = .880$, $p < .001$, $R^2 = .765$. This was also the case within each group: TA, standardized $\beta = .520$, $p < .001$, $R^2 = .270$; LA, standardized $\beta = .502$, $p < .001$, $R^2 = 0.252$.

Strategy use and problem size. As problem size increased, the participants were less likely to rely on retrieval and more likely to rely on counting (standardized $\beta = -.744$, $p < .001$, $R^2 = 0.554$). Within groups, both typically achieving and low achieving children showed the same pattern; TA, standardized $\beta = -.882$, $p < .001$, $R^2 = 0.778$; LA, standardized $\beta = -.736$, $p = .001$, $R^2 = 0.542$; Figure 6**.**

Relationship between strategy use and specific components of working memory. Across groups, regression analyses were used to examine the relation between performance on the strategy choice assessments and scores on the four

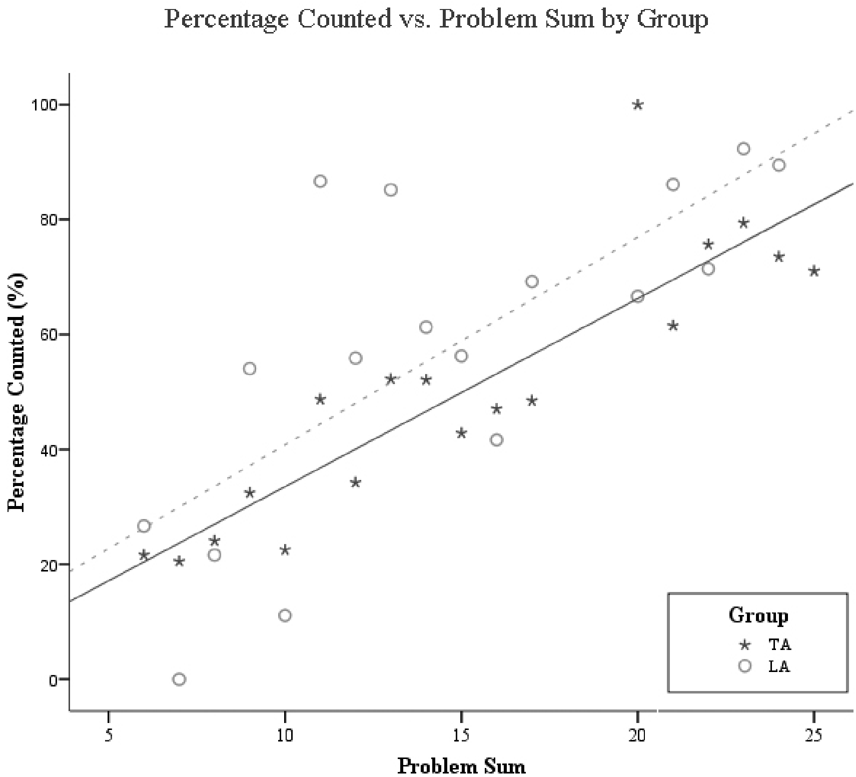

FIGURE 6 Both typically achieving (TA) and low achieving (LA) participants were more likely to use counting methods as problem sizes increased.

components of working memory (Table 2). Across groups, children with higher Counting Recall scores were more likely to retrieve, $F(1, 56) = 6.035$, $p = .015$, $R^2 = .101$. However, this relation did not hold within each group; TA, $F(1, 36) = 1.633$, $p = .210$, $R^2 = .043$; LA, $F(1, 18) = 3.831$, $p = .066$, $R^2 = .175$; Tables 3 and 4. A difference of slopes test indicated that these relationships between Counting Recall and retrieval were not significantly different between the two groups, $t(1) = 1.179$, $p = .244$.

Retrieved versus counted trials. Across groups, Backward Recall was inversely correlated with retrieved trial RT ($\beta = -.366$, $p < .01$, $R^2 = 0.135$), and Counting Recall was positively correlated with correct counting ($\beta = .325$, $p < .05$, $R^2 = 0.106$), and overall accuracy ($\beta = .270$, $p < .05$, $R^2 = 0.078$; Table 3). For the TA group, Backward Recall was inversely correlated with retrieved RT ($\beta = -.329$,

TABLE 2
Correlations Between WASI, WIAT, WMTB-C, and Strategy Assessment for All Participants

	WIAT			*WMBT-C*				*Strategy Assessment*	
	Numerical Operations	*Math Reasoning*	*Math Composite*	*Digit Recall*	*Backwards Recall*	*Counting Recall*	*Block Recall*	*% Retrieved*	*Response Times*
WIAT									
Numerical Operations	1.00	—	—	—	—	—	—	—	—
Math Reasoning	0.42***	1.00	—	—	—	—	—	—	—
Composite Score	0.88***	0.79***	1.00	—	—	—	—	—	—
WMBTC									
Digit Recall	–0.04	0.22	0.08	1.00	—	—	—	—	—
Backwards Recall	0.38**	0.23	0.36**	0.01	1.00	—	—	—	—
Counting Recall	0.05	0.19	0.13	0.26*	0.20	1.00	—	—	—
Block Recall	0.16	0.14	0.17	0.33**	0.24	0.27*	1.00	—	—
Strategy Assessment									
% Retrieved	0.03	0.10	0.08	–0.16	0.02	0.32*	–0.05	1.00	—
Response Times	–0.23	–0.01	–0.15	0.15	–0.17	–0.01	–0.04	–0.15	1.00

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

TABLE 3
Correlations Between WASI, WIAT, WMTB-C, and Strategy Assessment for the Typically Achieving Group

	WIAT			*WMBT-C*				*Strategy Assessment*	
	Numerical Operations	*Math Reasoning*	*Math Composite*	*Digit Recall*	*Backwards Recall*	*Counting Recall*	*Block Recall*	*% Retrieved*	*Response Times*
WIAT									
Numerical Operations	1.00	—	—	—	—	—	—	—	—
Math Reasoning	<u>0.24</u>	1.00	—	—	—	—	—	—	—
Composite Scale	0.77***	0.79***	1.00	—	—	—	—	—	—
WMBTC									
Digit Recall	<u>–0.09</u>	0.212	0.09	1.00	—	—	—	—	—
Backwards Recall	0.27	0.10	<u>0.21</u>	–0.11	1.00	—	—	—	—
Counting Recall	–0.29	0.04	–0.17	0.44**	0.14	1.00	—	—	—
Block Recall	0.19	–0.01	0.11	0.39*	0.21	<u>0.30</u>	1.00	—	—
Strategy Assessment									
% Retrieved	–0.10	0.11	0.03	–0.12	0.06	<u>0.21</u>	0.11	1.00	—
Response Times	0.10	0.07	0.10	0.23	–0.18	0.18	0.19	–0.22	1.00

Note: Underlined numbers correspond to correlations that were no longer significantly correlated once participants were broken into Typically Achieving and Low Achieving groups.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

TABLE 4
Correlations Between WASI, WIAT, WMTB-C, and Strategy Assessment for the Low Achieving Group

	WIAT			*WMBT-C*				*Strategy Assessment*	
	Numerical Operations	*Math Reasoning*	*Math Composite*	*Digit Recall*	*Backwards Recall*	*Counting Recall*	*Block Recall*	*% Retrieved*	*Response Times*
WIAT									
Numerical Operations	1.00	—	—	—	—	—	—	—	—
Math Reasoning	0.42*	1.00	—	—	—	—	—	—	—
Composite Score	0.86***	0.83***	1.00						
WMBTC									
Digit Recall	0.06	0.28	0.18	1.00	—	—	—	—	—
Backwards Recall	0.27	0.28	0.32	0.24	1.00	—	—	—	—
Counting Recall	–0.16	0.19	–0.01	0.10	0.06	1.00	—	—	—
Block Recall	0.31	0.40	0.40	0.25	0.34	0.29	1.00	—	—
Strategy Assessment									
% Retrieved	–0.39	–0.19	–0.36	–0.19	–0.34	0.42	0.10	1.00	—
Response Times	–0.27	0.08	–0.12	–0.04	0.03	–0.03	–0.42	0.30	1.00

Note. Underlined numbers correspond to correlations that were no longer significantly correlated once participants were broken into Typically Achieving and Low Achieving groups.

$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

$p < .05$, $R^2 = 0.108$), but no significant relations emerged for the LA group (Table 4).

DISCUSSION

Concordance Between Self- and Observer-Reports

The high agreement between the child and observer reports of strategy use suggests that children in the second and third grade are aware of the strategies they use to solve arithmetic problems, and that these self responses often concur with the overt signs noted by observers. Agreements between the participants' and the observers' reports within the TA and LA groups were virtually identical, suggesting that both groups of children are aware of the strategies they use. These findings further indicate that any differences in actual strategy use between the two groups are not the result of group differences in subjective reporting of strategy use.

ROC Analysis of Strategy Use

As predicted, retrieval trials were solved significantly faster than counting trials, reflecting the well-established differences in processing speed between an effortful production of the answer versus more rapid retrieval from memory (Siegler, 1988). Although this suggests a relationship between retrieval use and RT, it does not provide information about how well RT can be used to infer strategy use. Here, for the first time, we used an ROC to investigate this question. We also investigated the validity of other methods used to categorize strategy use.

As previously discussed, ROC curves can be used to test the appropriateness of a measure or metric for classifying data into groups (Centor, 1991). The results of our ROC analysis suggests that an ROC curve can appropriately categorize the strategy used to solve the problem via RT. For children in second and third grade, a cutoff threshold of 3662.5 ms maximized the sensitivity and specificity of the curve, indicating that trials with RTs greater than the threshold are most likely counting trials, whereas those below are retrieval trials. With this threshold, the resultant sensitivity and specificity indicated that the ROC curve was able to accurately classify a large proportion of the counting and retrieval trials; 78.7% of the counting trials were correctly classified, whereas 72.1% of the retrieval trials were correctly classified.

Depending on whether researchers choose to maximize the sensitivity or specificity of the ROC curve, the cutoff threshold will change accordingly. Maximizing the sensitivity would increase the number of counting trials that were correctly categorized, but would also risk incorrectly categorizing more retrieval trials. On the other hand, maximizing the specificity would decrease the number of counting trials that were incorrectly categorized, but at the risk of fewer correct categorization of retrieval trials. The determination of RT cutoff thresholds that maximizes the differentiability between the two strategies of counting and retrieval is particularly important for the efficient design of functional brain imaging studies (Menon, Meyer, & Wu, in press).

Interestingly, no differences were found between the ROC curves of the second and third graders. In agreement with this, there were no differences in math performance, working memory or strategy use. These findings suggest that significantly detectable shifts in strategy use may not occur between the second and third grades. Also, as noted by Siegel and Ryan (1989) and Siegler and Jenkins (1989), the shift in strategy choice is a gradual one, with children beginning to adopt retrieval while still using counting methods. Studies of changes in strategy choice across the school year confirm a gradual shift, at least for children in the United States (Geary, Bow-Thomas, Liu, & Siegler, 1996). Our data also point to considerable individual subject variability, which may have precluded detection of any overall differences. An interesting question is whether changes in the relationship between working memory and strategy use shifted between first and third grades found by Geary et al. (2004) reflect cognitive changes that are not captured by strategy use. In any case, these results point to the need for longitudinal studies to continually trace the development of strategy use between the first through third grades.

ROC analyses were also used to compare profiles of strategy use between the TA and LA groups. This analysis showed that the TA group had significantly higher AUC than the LA group, suggesting that RTs are related to strategy use with a higher sensitivity and specificity in the TA group. Our findings also suggest that our methods of strategy assessment are more optimal for detecting strategy use in typically achieving children. The differences in the ROCs between the two groups may be related to inconsistent strategy use and more variable performance in the LA group. For example, the LA group had slightly larger standard deviations in RT than the typically achieving children. Because of the variability in the response profile of counted and retrieved trials, the ROC analyses were less able to distinctly and accurately categorize strategy choices, as indicated by the lower peak sensitivity and specificity (True Positives and 1 - False Positives, respectively) for the LA group. Furthermore, the threshold for optimal sensitivity and specificity was also approximately 700 ms greater for the LA group than the TA group, suggesting that the low achieving children were slower overall in using both strategies, and most particularly when using the retrieval strategy. The different ROC profiles between the two groups are consistent with previous data which have suggested that children with MLD tend to lag behind typically achieving children in the development of strategy choice (Geary et al., 2007; Siegler, 1988) by at least one year (Geary et al., 2004). Our findings add to the growing evidence that many low achieving children have difficulty with math in part because they continue to apply less mature strategies and do not progress to the more efficient memory retrieval strategy (Geary et al., 2004).

Taken together, these ROC analyses strongly suggest that the reaction time differences reflect actual strategy choices and that the differences are more significant in typically achieving children than low achieving children. For the entire group, these findings suggest that a predictive model of strategy use can be derived from RT. Additional predictive models that are created from data sets on participants with other developmental patterns, problem types, can inform us greatly as to whether strategy choices can be predicted from behavior. Most importantly, this analysis provides an additional means of assigning strategy choice, and one that can be used concomitantly with other methods.

Strategy Use and Behavioral Performance

For the entire group, our analysis confirmed the recent findings of Geary et al. (2004) and earlier findings of Siegler (1987), in which RTs were greater when counting was used and in which RTs increased linearly with problem size. Overall, participants were more likely to use retrieval on the smaller problems, and use counting on larger problems.

Interestingly, there were no differences in the proportions of counters and retrievers between typically and low achieving children. Whereas low achieving children did not differ significantly from typically achieving children in the mix of strategies used, they were slower and had more performance errors. Our data support the view that low achieving children experience a persistent difficulty in retrieving arithmetic facts from long-term memory (LTM; Barrouillet & Lepine, 2005; Geary et al., 2004). Typically achieving children were more efficient in the use of their strategies, responding with faster RTs and greater accuracy when solving mental arithmetic problems as compared to low achieving children. Within the TA group, children showed faster RTs and were more accurate when employing the retrieval strategy compared to counting. The typically achieving children were also more accurate in counting than the low achieving children. Typically achieving children who retrieved may have been faster and more accurate due to their direct retrieval of correctly stored arithmetic facts in LTM. There were no significant differences within the two groups in accuracy on trials during which they counted versus those they retrieved. Our data suggest that low achieving participants were simply less proficient than typically achieving children regardless of the strategy they used.

In contrast to typically achieving children, low achieving children did not exhibit the profile of increased speed when they switched to the retrieval strategy. This finding suggests that they failed to show improvements in speed during the use of a method that is typically more efficient. Alternatively, they may have been implicitly counting on these trials. Although low achieving children were spending more time than typically achieving children overall, they remained less accurate and the additional time used to count did not translate to any accuracy differences. This result is consistent with previous findings that low achieving children and children with MLD commit more errors when either finger counting or retrieval is used when solving both simple and complex problems (Geary et al., 2004). It is therefore likely that even when children with MLD eventually shift to retrieval of arithmetic facts from LTM, their processing is less efficient, perhaps due to weaker working memory (Barrouillet & Lepine, 2005).

Strategy Use, Performance, and Role of Specific Working Memory Components

Our findings highlight the importance of the central executive, rather than the phonological loop and the visuospatial sketchpad, as the key component of working memory that is related to strategy use. With the exception of Geary et al., (2007) and other reported in this special issue, little research has investigated how each of the specialized components of working memory differentially contributes to strategy use in children. For the entire group of participants in the present study, Counting Recall, a measure of the central executive component of working memory, was positively correlated with the percentage of trials in which each child used retrieval. This result is in agreement with the finding that as counting span performance increases in first graders, reliance on finger counting and finger counting errors decreases (Geary et al., 2004). Our findings complement these results by suggesting that these relations extend beyond first grade to children in the second and third grades. An intriguing question for further research is how these processes might change between the first and second grade, particularly in view of Passolunghi and colleague's findings of changes in the contribution of working memory and general math performance during this time period (2008/this issue).

Second, our findings clarify the role of all three components of working memory in strategy use. Three previous studies (Barrouillet & Lepine, 2005; Geary et al., 2004; Geary et al., 2007) have examined the role of working memory in strategy choice using measures associated with the central executive. The procedure used in the Geary et al. (2004) study, based on Hitch and McAuley (1991), is similar to the Counting Recall task in the *WMBT-C*. Briefly, children are presented with target red and distractor blue dots randomly placed on index cards and were asked to count, remember, and then recall the number of red dots on each card. Barrouillet and Lepine used a similar counting task with target and distracter dots that were interspersed with letters to be remembered. Given the similarities in the counting tasks used in the previous and present studies, the findings seem to converge on the central executive as the specific component of working memory that mediates strategy choice. Our study builds on these previous studies by also including another standardized measure of the central executive (i.e., Backwards Recall), and standardized measures of the phonological loop and the visual spatial sketchpad. A novel finding of our study is that these measures were not significantly correlated with retrieval use or with RTs, suggesting that the central executive may play the main role in orchestrating simple fact retrieval in second and third graders.

Our results also provide new insights into the influence of working memory on the counted and retrieved trials, as assessed and validated by our measures of strategy use. The central executive influenced performance on both types of trials, but in slightly different ways. Counting Recall was related to accurate task performance on the counted trials, whereas Backwards Recall was related to the rapidity with which retrieval was used. These findings indicate that, within the same group of children, the central executive not only contributes to flexible strategy use, but also to efficient task performance on the counted and retrieved trials. The phonological loop and visuospatial sketchpad were not related to strategy use on either strategy, but Geary and colleagues (2007) did find that these components of working memory were important for other areas of early mathematics learning.

Typically achieving children, compared to low achieving children, performed significantly better on Backward and Counting Recall tasks of the central executive component of working memory. No significant differences were found between the two groups for the phonological loop and visuospatial sketchpad capacity. These findings concur with previous findings that demonstrate a persistent working memory deficit in children with poor mathematics achievement scores (D’Amico & Guarnera, 2005; McLean & Hitch, 1999). Geary and colleagues (2004) found that in a sample of first graders, working memory capacity mediated the greater use of verbal counting by the normally achieving children and contributed to the greater use of finger counting by the children in their MLD group. Consistent with these findings, typically achieving children in the current study tended to retrieve slightly more and were faster in responding during the strategy assessment.

Although the IQ scores of typically achieving children were not significantly different from low achieving children, the typically achieving children demonstrated higher working memory capacities and significantly better mathematical achievement (as indexed by the WIAT—II) compared to the low achieving group. Thus, differences in working memory and mathematical achievement most likely reflect a domain-specific, developmental delay, as opposed to fixed disparities in aptitude. Indeed, Geary and colleagues (2004) found that children with MLD present strategy choice and working memory profiles typically one year behind those of their typically achieving peers. This finding has implications for the development of curricula geared towards children with MLD, and perhaps their low achieving peers. Further research is needed to determine whether this delay in children with MLD can be improved with intervention geared toward greater use of retrieval strategies (Fuchs, Fuchs, & Karns, 2001).

The influence of working memory on performance also differed between children in the TA and LA groups. Children in the TA group exhibited the same behavior as the entire group, wherein Backwards Recall was correlated with increased efficient retrieval. Children in the LA group showed no such effects. As noted, compared to typically achieving children, low achieving children demonstrated significantly lower scores on the Backward Digit and Counting Recall tasks of the central executive. The central executive component of working memory may orchestrate the storage of arithmetic facts calculated in working memory into LTM. On the other hand, low achieving children may not experience difficulty in storing a fact in long-term memory, but instead difficulty in retrieving a stored fact. In this case, the episodic buffer, which is thought to bind representations from a subsystem and LTM into a unitary episodic representation (Baddeley, 2000) may not function adequately in low achieving children. Further research is needed to investigate how the central executive influences storage and consolidation in, and retrieval of facts from, LTM.

To summarize, low achieving children had lower scores on their central executive and phonological loop measures, despite high average IQ scores, and showed lower accuracy and slower RTs when they used retrieval strategies. In contrast with typically achieving children, working memory components in low achieving children do not play a significant role in efficient retrieval. Developmental delays in the central executive in low achieving children may therefore contribute to their poor accuracy and slow RT when retrieving basic arithmetic facts from LTM.

CONCLUSION

ROC analysis provides new and more thorough validation of strategy use as developed by Siegler, Geary, and others. Standardizing strategy choice assessment in second- and third-grade children is useful for discerning strategy maturation as well as for isolating strategy deficits in low achieving children. ROC analysis may be particularly helpful for designing brain imaging experiments based on individual differences in strategy use. Our study further emphasizes the role of the central executive, rather than the phonological loop and visuospatial sketchpad, component of working memory as a key cognitive factor that mediates efficient strategy selection and use. Although the proportion of counters and retrievers were analogous in the two groups, low achieving children take more time and commit more errors than the typically achieving children. Our findings also suggest that the central executive influences performance on trials regardless of whether retrieval or counting strategies are used. Together with the observation that low achieving children scored significantly lower on standardized measures of the central executive, these results suggest that the Central Executive component of working memory plays an important role in both the choice and execution of particular strategies.

Our study has focused on the TA and LA groups. It is important to note that our findings regarding the LA group may not necessarily generalize to children who are strictly MLD. Further research with three well-separated groups will be necessary to clarify this important issue.

ACKNOWLEDGMENTS

We thank Katherine Keller, Jose Anguiano, and Sabina Lau for their assistance with data acquisition and scoring. We also thank Kaustubh Supekar for his assistance with the ROC curve methods. This research was supported by NIH grant HD047520.

REFERENCES

Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review, 2*, 213–236.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*, 75–106.

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

Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology, 33*, 216–234.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A*, 5–28.

Baddeley, A. (2000). The episodic buffer: A new component of WM? *Trends in Cognitive Science, 4*, 417–423.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.

Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology, 51A*, 819–852.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory* (vol. 8, pp. 47–89). New York: Academic Press.

Baroody, A. J. (1987). *Children's mathematical thinking. A developmental framework for preschool, primary, and special education teachers*. New York: Teachers College Press.

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

Barrouillet, P., & Lepine, R. (2005). WM and children's use of retrieval to solve addition problems. *Journal of Experimental Child Psychology, 91*, 183–204.

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.

Carpenter, T. P. M., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics and Education, 15*, 179–202.

Centor, R. M. (1991). Signal detestability: The use of ROC curves and their analyses. *Medical Decision Making, 11*, 102–106.

Cooney, J. B., Swanson, L. H., & Ladd, S. F. (1988). Acquisition of mental multiplication skill: Evidence for the transition between counting and retrieval strategies. *Cognition and Instruction, 5*, 323–345.

D'Amico, A., & Guarnera, M. (2005). Exploring WM in children with low arithmetical achievement. *Learning and Individual Differences, 15*, 189–202.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia, 37*, 1189–1199.

Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartner's mathematical development: Effects of peer-assisted learning strategies. *The Elementary School Journal, 101*, 495–510.

Fuson, K. C. (1982). An analysis of the counting-on solution procedure in addition. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (eds.), *Addition and subtraction: A cognitive perspective* (pp. 67–81). Hillsdale, NJ: Lawrence Erlbaum.

Gathercole, S. E., & Adams, A. (1994). Children's phonological WM: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language, 33*, 672–688.

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., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67*, 2022–2044.

Geary, D. C., & Brown, S. C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology, 27*, 398–406.

Geary, D. C., & Burlingham-Dubree, M. (1989). External validation of the strategy choice model for addition. *Journal of Experimental Child Psychology, 47*, 175–192.

Geary, D. C., & Widaman, K. F. (1987). Individual differences in cognitive arithmetic. *Journal of Experimental Psychology: General, 116*, 154–171.

Geary, D. C., Bow-Thomas, C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67*, 2022–2044.

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 WM and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88*, 121–151.

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2007). Strategy use, long-term memory, and WM capacity. In D. B. Berch, & M. M. Mazzocco (eds.), *Why is math so hard for some children?* (pp. 83–105). Baltimore, MD: Paul H. Brookes Publishing Co.

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., Widaman, K. F., Little, T. D., & Cormier, P. (1987). Cognitive addition: Comparison of learning-disabled and academically normal elementary school children. *Cognitive Development, 2*, 249–269.

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

Hanich, L., Jordan, N., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning disabilities. *Journal Educational Psychology, 93*, 615–626.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29–36.

Hitch, J. G. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology, 10*, 302–323.

Hitch, J. G., & McAuley, E. (1991). WM in children with specific arithmetical learning difficulties. *British Journal of Psychology, 82*, 375–386.

Houlihan, D. M., & Ginsburg, H. P. (1981). The addition methods of first- and second-grade children. *Journal for Research in Mathematics Education, 12*, 95–106.

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

Jordan, N., Hanich, L., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*, 834–850.

Katz, D., & Foxman, B. (1993). How well do prediction equations predict? Using receiver operating characteristic curves and accuracy curves to compare validity and generalizability. *Epidemiology, 4,* 319–326.

Kaye, D. B., Post, T. A., Hall, V. C., & Dineen, J. T. (1986). Emergence of information-retrieval strategies in numerical cognition: A developmental study. *Cognition and Instruction, 3*, 127–150.

Logie, R. H. (1986) . Visuo-spatial processing in working memory. *Quarterly Journal of Experimental Psychology, 38A*, 229–247.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition, 22*, 395–410.

Lusted, L. B. (1971). Signal detectability and medical decision-making. *Science: New Series, 171*, 1217–1219.

Mazzocco, M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research and Practice, 20*, 142–155.

McLean, J. F., & Hitch, G. J. (1999). WM impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology. Special Issue. The Development of Mathematical Cognition: Arithmetic, 74*, 240–260.

Menon, V., Meyer, M., & Wu, S. (in press). Cognitive developmental neuroscience of mental arithmetic: Implications for understanding learning and academic achievement. In A. E. Kelly (ed.), *Neuroscience and Mathematics Education.* Cambridge University Press.

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.

Passolunghi, M. C., Mammarella, I. C., & Altoè, G. (2008/this issue). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Developmental Neuropsychology, 33,* 229–250.

Pickering, S. J., & Gathercole, S. E. (2001). *WM Test Battery for Children (WMTB-C) Manual*. London: The Psychological Corporation.

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*, 137–157.

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

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*, 405–410.

Siegel, L. S., & Ryan, E. B. (1989). The development of WM in normally achieving and subtypes of learning disabled children. *Child Development, 60*, 973–980.

Siegler, R. S. (1986). *Unities across domains in children's strategy choices*. Hillsdale, NJ: Erlbaum.

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., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Robinson, M. (1982). The development of numerical understanding. *Advances in Child Development and Behavior, 16*, 241–312.

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). Hillside, NJ: Erlbaum.

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

Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology, 98*, 265–281.

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.

Svenson, O. (1985). Memory retrieval of answers of simple additions as reflected in response latencies. *Acta Psychologica, 59*, 285–304.

Weschler, D. (2005). *Wechsler Individual Achievement Test* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Weschler, D. (1999). *Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: Harcourt Assessment, Inc.

Widaman, K. F., Little, T. D., Geary, D. C., & Cormier, P. (1992). Individual differences in the development of skill in mental addition: Internal and external validation of chronometric models. *Learning & Individual Differences, 4*, 167–213.

Wilson, K. M., & Swanson, H. L. (2001). Are mathematics disabilities due to a domain-general or a domain-specific working memory deficit? *Journal of Learning Disabilities, 34*, 237– 248.