



Published in final edited form as:

J Psychoeduc Assess. 2009 June ; 27(3): 265–279. doi:10.1177/0734282908330592.

Predicting Mathematical Achievement and Mathematical Learning Disability With a Simple Screening Tool: The Number Sets Test

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

University of Missouri

Abstract

The Number Sets Test was developed to assess the speed and accuracy with which children can identify and process quantities represented by Arabic numerals and object sets. The utility of this test for predicting mathematics achievement and risk for mathematical learning disability (MLD) was assessed for a sample of 223 children. A signal detection analysis of first grade Number Sets Test scores provided measures of children's sensitivity to number and their response bias. The sensitivity measure, d' , but not the response bias measure was predictive of third grade mathematics achievement scores, above and beyond the influence of intelligence, working memory, and first grade achievement scores. Further analyses assessed the sensitivity and specificity of the test and revealed that first grade d' scores identified 2 out of 3 children diagnosed as MLD in third grade and correctly identified about 9 out of 10 children who were not at risk for MLD.

Keywords

Mathematics; Learning Disability; Number Sense; Screening Test

Between 5% and 10% of children will be diagnosed with some form of learning disability in mathematics (MLD) by the time they complete high school (Barbarese, Katusic, Colligan, Weaver, & Jacobsen, 2005; Gross-Tsur, Manor, & Shalev, 1996; Kosc, 1974; Ostad, 1998; Shalev, Manor, & Gross-Tsur, 2005). The early identification of risk for MLD and thus potential for early intervention is critical on several levels. Once children are behind in mathematics, they tend to stay behind (Shalev et al., 2005). Early mathematical skills are also an important predictor of later educational achievement in general. On the basis of a meta-analysis of six longitudinal studies and controlling for socioemotional status, cognition, and attention, as well as family background, Duncan et al. (2007) showed that the strongest predictor for later achievement in mathematics and reading was school-entry mathematics skills.

Although the specific cognitive systems that contribute to MLD and the deficits that define MLD in young children are still under investigation, certain areas have emerged as playing key roles. In comparison to normally achieving (NA) children, children with MLD use less mature, inefficient counting strategies when solving arithmetic problems; they have a poor conceptual understanding of counting; and they have persistent difficulty learning and retrieving basic arithmetic facts from long-term memory (Geary, 1993, 2004; Jordan, Hanich, & Kaplan, 2003; Ostad, 1998). Of the underlying cognitive systems, the central executive component of working memory has emerged as a contributing factor in many of these areas (Barrouillet, Fayol, & Lathulière, 1997; Geary, Hamson, & Hoard, 2000; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007). These studies have provided insights into the deficits of children with MLD when learning school mathematics, but the early identification of MLD will need to focus on the core competencies that define quantitative knowledge before school entry (Gersten, Jordan, & Flojo, 2005).

Number sense is one such entry-level competency. Although disagreements remain as to the definition and scope of number sense, at the very least it involves children's implicit understanding of the absolute and relative magnitude of sets of objects and of symbols (e.g., Arabic numerals) that represent the quantity of these sets. Children's early number sense will manifest in their ability to immediately identify the numerical value associated with small quantities, a facility with use of counting to quantify small sets of objects and to add and subtract small quantities to and from these sets, and a proficiency in approximating the magnitudes of small numbers of objects and simple numerical operations (Dehaene, 1997; National Mathematics Advisory Panel, 2008). This intuitive sense of quantity and magnitude may be inherent (Butterworth & Reigosa, 2007; Dehaene, Piazza, Pinel, & Cohen, 2003) and may provide the foundation for early mathematics learning in school (Geary, 2006). Moreover, preliminary evidence suggests that children with MLD have a deficit in this fundamental understanding of number and magnitude (Koontz & Berch, 1996; Landerl, Bevan, & Butterworth, 2003).

We designed the Number Sets Test as a group-administered pencil-and-paper measure of the speed and accuracy with which children can identify number and quantity of sets of objects and combine these with quantities represented by Arabic numerals. The combination thus potentially taps critical features of number sense (Geary et al., 2007). Our first study indicated that children with MLD have a deficit on the competencies assessed by this measure that is unrelated to intelligence (IQ) and that is above and beyond their deficit on the central executive. The current study extends these findings in two ways. The first was an exploration of the component skills assessed by the Number Sets Test and the second an assessment of the potential utility of this test as a quick (about 10 min) screening measure of risk for MLD. The component skills potentially contributing to performance on this test include IQ, working memory, and competencies in number, counting, and arithmetic.

The specific quantitative measures from Geary et al. (2007) that are used in this study include the percentage of addition facts correctly retrieved from long-term memory, the detection of double counting errors on a counting knowledge task, and the accuracy (i.e., degree of error) of the child's placement on a number line task. Performance on each of these measures is predictive of later mathematics achievement and MLD and is influenced by schooling (e.g., Geary, Bow-Thomas, & Yao, 1992; Jordan et al., 2003; Siegler & Booth, 2004). Nevertheless, the measures may also capture components of children's early number sense. As an example, performance on number line task is dependent on the same area of the parietal cortex that supports processing of magnitude and general quantity (Zorzi, Priftis, & Umiltà, 2002). If the Number Sets Test taps these same competencies, then we should find a substantive correlation between performance on the Number Sets Test and the number line task. In any case, the diagnostic utility of the Number Sets Test was assessed by determining the sensitivity and specificity of first grade scores for predicting MLD status at the end of third grade.

Method

Participants

All kindergarten children from 12 elementary schools were invited to participate in a longitudinal prospective study of MLD. Parental consent and child assent were received for 37% ($n = 311$) of these children (see Geary et al., 2007). The current analyses include the 228 children (105 male) with kindergarten, first, second, and third grade achievement scores and first grade IQ, working memory, and mathematical cognition test scores.

Standardized Measures

Intelligence—In the first grade, the children were administered one verbal (Vocabulary) and one nonverbal (Matrix Reasoning) subtest of the Wechsler Abbreviated Intelligence Scale (WASI; Wechsler, 1999), and these scores were used to estimate IQ based on norms presented in the manual. For the age range of our sample, the manual reports a correlation of .92 to .93 between IQ based on all four subtests of this scale and IQ based on the two subtests we used.

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

Mathematical Tasks

Number sets test—The child's task was to determine as quickly and accurately as possible if pairs or trios of object sets, Arabic numerals, or a combination of these matched a target number (5 and 9); the items of 5 and 9 were chosen because they represent smaller and larger values within the range of basic Arabic numerals (i.e., 1 to 9). As shown in Figure 1, the object sets or numerals were combined to create domino-like rectangles; specifically, two types of stimuli were developed: 0 to 9 small objects (circles, triangles, diamonds, and stars) in a ½-inch square and one Arabic numeral (18 point font) in a ½-inch square. Each test page also includes two lines of three three-square rectangles for each combination. The target numbers were listed in a large font (36 point) at the top of each page. On each page, 18 items matched the target, 12 were larger than the target, 6 were smaller than the target, and 6 contained 0 squares or an empty square.

Two items matching a target number of 4 were first explained for practice. Using 3 as the target number, four lines of two items were then administered as practice. Once it was determined that the child understood the task, the experimental items were administered. The child was instructed to move across each line of the page from left to right without skipping any and 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.” Using a stopwatch, the child was given 60 s and 90 s per page for the targets 5 and 9, respectively, and was asked to stop at the time limit. We chose to time the task to avoid ceiling effects and because a timed measure should provide an assessment of fluency in recognizing number combinations. The task yields information on the number of items correctly identified (i.e., circled) as matching the target value, hits (α , $\alpha = .88$); the number of correct matches that were not identified, misses ($\alpha = .70$); the number of incorrect items that were not circled and thus rejected as matches, correct rejections (CRs; $\alpha = .85$); and the number of incorrect items that were identified as matching the target, false alarms (FAs; $\alpha = .90$). Alpha values are from Geary et al. (2007).

Number 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 and Booth (2004). 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. The overall score is the mean of these differences across trials.

Counting knowledge—The child watched a puppet count a series of alternating red and blue chips at several different counting string lengths. Sometimes the puppet counted correctly and other times did not. The correct counts could be the standard left to right count or could be nonstandard (e.g., right to left). The incorrect counts violated a basic counting principle, such as one-one correspondence (Gelman & Gallistel, 1978). The child was asked to state if the way the puppet counted was “OK” or “not OK”, and thus, we assessed the child’s awareness of counting principles and his or her understanding of whether variation from the standard left to right counting could still be correct (Briars & Siegler, 1984). Across a series of studies, we have found that children with MLD consistently miss trials on which the first chip in the sequence is double counted (e.g., Geary et al., 1992). Therefore, the variable used in this study was the percentage of trials (out of three) on which the child correctly detected this counting error.

Addition strategy assessment—Fourteen simple (e.g., $3 + 6$) and six complex ($9 + 15$) addition problems were horizontally presented in a large font (about 2 cm tall), one at a time, at the center of a 5-inch by 8-inch card (see Geary et al., 2007). The child was asked to solve each problem (without the use of 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 the child was instructed to speak the answer out loud. Based on the child’s answer and the experimenter’s observations, the trial was classified into one of six strategies—specifically, counting fingers, fingers, verbal counting, retrieval, decomposition, or other/mixed strategy (Siegler & Shrager, 1984). A mixed trial was one in which the child started using one strategy but completed the problem using another strategy. Counting trials were further classified as min. (i.e., the larger addend was stated and the smaller addend was counted, as in counting 3, 4, 5 to solve $3 + 2$), sum (both addends were counted), or max. (the smaller addend was stated and the larger was counted).

During problem solving, the experimenter watched for physical indications of counting, such as regular finger (finger counting) or mouth (verbal counting) movements. On verbal counting trials, the experimenter probed the child as to how she counted, and the child’s response was recorded. If the child held out a number of fingers to represent the addends and then stated an answer without counting them, then the trial was initially classified as fingers. If the child spoke the answer quickly, without hesitation, and without obvious counting-related movements, then the trial was initially classified as direct retrieval or as decomposition if this was the child’s predominant retrieval-based strategy on previous trials; decomposition involves, as an example, solving $7 + 8$ by decomposing 8 into 5 and 3 and then adding $7 + 3 = 10$, $10 + 5 = 15$. After the child had spoken the answer, the experimenter queried the child on how the answer was obtained. If the child’s response (e.g., “just knew it”) differed from the experimenter’s observations (e.g., saw the child mouthing counting), then a notation indicating disagreement between the child and the experimenter was made. If counting was overt, then the experimenter classified it as a counting strategy. If the trial was ambiguous, then the child’s response was recorded as the strategy. Previous studies indicate that this method provides a useful measure of children’s trial-by-trial strategy choices (e.g., Siegler, 1987). The percentage correct direct-retrieval trials for simple problems—direct retrieval was uncommon for the complex problems—is correlated with mathematics achievement scores (Geary, Bow-Thomas, Liu, & Siegler, 1996) and is an indicator of MLD (Geary, 2004). We used this percentage in the current analysis.

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 sketchpad. All of the subtests have six items at span levels ranging from one to six to one to

nine. Passing four items at a level moves the child to the next level. At each span level, the number of items (e.g., words) to be remembered is increased by one. Failing three items terminates the subtest.

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 four, five, six, or seven 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 variant 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 only be seen 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.

Procedure

All children were tested in the spring of their kindergarten year and in the fall and spring of first, second, and third grade. The spring assessments included the achievement tests and the fall assessments the mathematical tasks. Because of testing time considerations, the number line test was moved to the spring assessment for second and third grade (see Geary et al., 2007). The majority of children were tested in a quiet location at their school site and occasionally in a testing room on the university campus or in a mobile testing van if the child moved between assessments. The WMTB-C was administered in the testing van or on the university campus during first grade.

Results

In the first section, we describe a signal detection analysis of performance on the Number Sets Test. The analysis enables separation of children's sensitivity to number from their tendency to respond (i.e., circle) to test items. We then examine the contributions of IQ, working memory, and performance on the mathematical cognition measures to children's sensitivity to number. In the second and third sections, we assess the utility of the Number Sets Test as a predictor of individual differences in mathematics achievement and as a potential screening measure for MLD.

Sensitivity to Number Sets

The design of the Number Sets Task allows for a signal detection analysis of children's accuracy (MacMillan, 2002). The key variables are sensitivity (d') and response bias (C). The former represents the child's sensitivity in the detection of target quantities (i.e., 5 or 9) and is calculated as the difference between the z -score for hits and FAs. On the basis of preliminary

analyses, d' was calculated using the total number of items, not simply the number of items attempted. The response bias represents the child's tendency to respond to task items, regardless of whether they are correct, and was calculated as the average of the z scores for hits and FAs—specifically, following MacMillan, $-.5(z \text{ hits} + z \text{ FAs})$. Children who correctly identify many target quantities and make few FAs will have high d' and low C scores, whereas children who have as many hits as FAs will have low d' and high C scores. In the latter case, the high number of correct items is due to the child's bias to respond and not sensitivity to quantity.

Higher d' scores were associated with higher mathematics achievement ($ps < .001$) in kindergarten ($r = .58$), first grade ($r = .50$), second grade ($r = .52$), and critically third grade ($r = .49$). The correlations between C scores and achievement were less consistent, and the critical correlation with third grade mathematics achievement scores ($r = -.14$) indicated little predictive value of C above and beyond d' . Also, d' scores were significantly ($ps < .001$) correlated with IQ ($r = .48$), each of the three components of working memory ($rs = .35$ to $.51$), number of addition facts correctly retrieved from long-term memory ($r = .37$), detection of counting errors ($r = .37$), and degree of error on the number line task ($r = -.61$).

Regressing z scores for d' values on z scores for these measures revealed significant and unique contributions from IQ and central executive scores, as well as significant contributions from each of the mathematical cognition measures, as shown in Table 1; the phonological loop and visuospatial sketchpad scores were not significant. Overall, these measures explained 52% of the variation in d' scores, $F(5, 222) = 48.3, p < .001$.

Prediction of Third Grade Achievement

All predictors in the following regressions were centered (to consider quadratic components and interactions) and standardized. The unstandardized beta weights represent the number of points increase in percentile rank in third grade achievement from one standard deviation increase in the predictor.

Mathematics—First grade mathematics achievement scores, IQ, the three working memory scores, and d' (all measured in first grade) were used to predict third grade mathematics achievement scores. Preliminary analyses revealed the phonological loop and visuospatial sketchpad were not unique predictors of third grade mathematics achievement and were thus dropped. Table 2 summarizes the results for the significant predictors; specifically, first grade mathematics achievement scores (percentile rank), IQ, central executive scores, and d' scores, $R^2 = .37, F(4, 223) = 32.73, p < .001$. Examination of the table reveals that central executive scores are the best predictor of third grade mathematics achievement, but first grade mathematics achievement scores, IQ, and d' all predict similar and unique amounts of variance.

To determine which single measure would be most useful in predicting third grade mathematics achievement, each of the four measures in Table 2 were run in independent regressions. The central executive and d' measures predicted 25%, $F(1, 226) = 74.00$, and 24%, $F(1, 226) = 72.18$, respectively, of the variation in third grade mathematics achievement scores ($ps < .001$). In comparison, first grade mathematics achievement scores and IQ predicted 16%, $F(1, 226) = 41.70$, and 18%, $F(1, 226) = 50.73$, respectively, of the variation in third grade mathematics achievement scores ($ps < .001$).

We also assessed the utility of number line scores (i.e., degree of deviation from the correct placement), percentage of correct direct retrieval on the addition strategy task, and counting error scores for the prediction of third grade mathematics achievement. In one predictor regression, the number line scores were a slightly better predictor than d' —explaining 27% of the variation in third grade mathematics achievement ($p < .001$)—and percentage correct

retrieval a slightly worse predictor—explaining 20% of the variation ($p < .001$). Counting error scores explained 7% of the variation in third grade mathematics achievement scores ($p < .001$).

Reading—To assess discriminant validity, first grade reading achievement scores, IQ, the three working memory scores, and d' were used to predict third grade reading achievement scores. Of the working memory measures, only the phonological loop predicted unique variation, and thus, the central executive and visuospatial sketchpad variables were dropped. Table 3 summarizes the results for first grade reading achievement scores (percentile rank), IQ, and phonological loop and d' scores, $R^2 = .72$, $F(4, 223) = 145.40$, $p < .001$. Examination of the table reveals that first grade reading achievement was the best predictor of third grade reading scores. In fact, only phonological loop scores added unique variance to the prediction above and beyond first grade scores. Critically, d' was not a unique predictor of third grade reading achievement.

Predicting MLD

Students were classified as having MLD if they scored below the 15th percentile on both the second and third grade mathematics achievement test ($n = 45$, 16 male). The use of cutoffs across two consecutive grades is useful for identifying students with cognitive deficits related to mathematics learning and who are thus at risk for continued poor achievement (see Geary et al., 2007; Murphy, Mazzocco, Hanich, & Early, 2007). Students who scored between the 15th and 30th percentiles, inclusive, in both second and third grade were classified as low achieving (LA; $n = 17$, 8 male) and students who scored above the 30th percentile in both grades as NA ($n = 96$, 58 male). Students who did not score in the same range in consecutive years were not included in this analysis. Mean achievement, IQ, working memory, and mathematical cognition z scores are shown in Table 4.

Corresponding categories were then created based on first grade mathematics achievement scores and d' scores. For example, students scoring below the 15th percentile on the first grade mathematics test or below the 15th percentile of our sample on the d' measure were initially diagnosed as MLD. Table 5 shows the frequency with which the initial, predicted category—MLD, LA, NA—based on first grade scores corresponded to the final diagnostic category—MLD, LA, NA—at the end of third grade. To illustrate, 23 children were initially predicted to be MLD based on their first grade mathematics achievement, and 18 of these children were diagnosed as MLD in third grade, and one and four, respectively, were classified as LA and NA. Of the 45 children diagnosed as MLD in third grade, 18, 19, and 8 were initially classified as MLD, LA, and NA, respectively.

The second set of values represents the number of children in the same group in both grades (hit), the number of children incorrectly diagnosed (FA), the number of children correctly rejected (CR) as not belonging in the third grade group, and the number of third grade children missed by the first grade grouping. Using an initial first grade diagnosis of MLD, 18 of these children were correctly categorized (hit) as MLD in third grade and 5 children were falsely identified as MLD; that is, they were in the LA ($n = 1$) or TA ($n = 4$) groups in third grade. Of the remaining children, 108 children were correctly rejected as not MLD in first grade, and 27 children were misses. The latter groups are children who were placed in the LA or NA groups in first grade but were diagnosed as MLD in third grade.

The diagnostic utility of tests is considered in terms of sensitivity and specificity (Altman & Bland, 1994). Sensitivity is the ratio of true positives to total positives, and specificity is the ratio of true negatives to total negatives. The top section of Table 6 shows these values using the 15th percentile cutoff for first grade mathematics achievement and d' scores for diagnosing third grade MLD. The specificity of both measures is high—96% of children who did not have MLD in third grade were correctly identified as non-MLD in first grade. The sensitivity—the

percentage of third grade children with MLD correctly diagnosed in first grade—of d' is higher than that of first grade mathematics achievement scores, although neither is particularly high.

Using the MLD and NA groups from the previous analysis, we used response operating curves (ROCs) to maximize sensitivity and specificity of the first grade mathematics achievement and d' scores in predicting third grade MLD. The maximized values are shown in the bottom section of Table 6. The sensitivity of both measures is improved, with a modest reduction in specificity for d' and a substantial reduction for first grade mathematics achievement scores. Figure 2 was created using software by Sing, Sander, Beerenwinkel, and Lengauer (2005) and is the ROC curve using d' to predict third grade MLD. The horizontal line is the sensitivity cutoff, and the vertical line is the specificity cutoff. The figure illustrates a sensitivity cutoff at 90%—that is, 90% of the third grade MLD group lies below this cutoff. The corresponding specificity is 56%; 56% of the non-MLD children are identified as such. When first grade mathematics achievement scores are set at a sensitivity of 90%, the corresponding specificity is 43%.

Discussion

On the basis of previous research on the cognitive mechanisms that may contribute to MLD, it has been hypothesized that these children have a fundamental deficit in the ability to represent and process representations of small quantities, a basic deficit in a core aspect of number sense (e.g., Butterworth & Reigosa, 2007; Geary et al., 2007; Koontz & Berch, 1996). The Number Sets Test was developed to provide an easy to administer—individually or in groups—and short (~10 min) pencil-and-paper measure of the speed and accuracy of processing Arabic and object-set representations of relatively small quantities. Our goals were to decompose the competencies that contribute to individual differences on this test and to assess its feasibility as a potential screening test for MLD.

Assessed Competencies

The use of signal detection methods allowed us to separate children's sensitivity to representations of number from their tendency to be conservative or lenient in their response bias (identify the items as matching the target or not). The finding that the d' measure correlated with later mathematics achievement but the C measure did not confirms the utility of this approach. The regression analyses revealed that children with higher IQ scores and higher executive scores had above average d' scores. This relation is not unusual, given that IQ and the components of working memory assessed by central executive measures are typically correlated with achievement and ability measures (e.g., Walberg, 1984). The important finding is that the number, counting, and addition measures were related to d' after controlling for IQ and the central executive.

In fact, the best predictor of d' scores was performance on the number line measure; more accurate number line placements were associated with higher d' scores. This number line measure is also correlated with mathematics achievement (Siegler & Booth, 2004) and appears to assess, at least in part, children's intuitive understanding of numerical quantity, a core aspect of number sense. The brain and cognitive systems that support numerical quantities as they might be represented on a number line are hypothesized to be the same as those that support an intuitive understanding of numbers represented by sets (Dehaene et al., 2003; Gallistel & Gelman, 1992), as in our task. Whether the relation between our d' scores and accuracy on the number line task is because both tests are dependent on these same brain and cognitive systems remains to be established. Our results, nevertheless, provide initial support for this hypothesis. Whatever competencies are being assessed by d' , they are unrelated to reading achievement. The d' measure predicts something unique to later mathematics achievement that is unrelated to achievement in general or to IQ or working memory.

Screening Measure

The Number Sets Test has promise as a potential screening tool for identifying children at risk for MLD. The d' scores based on administration of the test in early first grade can be used to identify two out of three children identified as having MLD at the end of third grade (sensitivity) and to correctly identify nearly 9 out of 10 non-MLD children (specificity). The corresponding ROC analysis also provides a curve that allows determination of the tradeoffs between sensitivity and specificity. The values reported in Table 6 are determined mathematically, but the tradeoffs may differ practically.

In terms of practice, the optimal cutoff values depend on the costs and benefits of early identification and remediation. If remediation is inexpensive and easily achieved (see Siegler & Ramani, 2008, for a promising approach), then maximizing sensitivity, regardless of changes in specificity, is likely to be the best approach. This is because the majority of children at risk for MLD will be identified and provided remedial services, and the costs of providing these services to children who do not need them is small. With limited resources, however, sensitivity must be balanced against specificity so that resources can be most effectively used. We caution that the Number Sets Test is not yet ready for use as a diagnostic instrument because it is not normed and because we do not yet have information on its predictive validity for grades beyond third grad. Nevertheless, to enable further tests of its utility, we will provide, on request, the test and a table for other researchers to translate the ROC curve into actual Number Sets Test scores.

Acknowledgments

Geary acknowledges support from Grants R01 HD38283 from the National Institute of Child Health and Human Development (NICHD) and R37 HD045914 cofunded by NICHD and the Office of Special Education and Rehabilitation Services. We thank Linda Coutts, Kendra Andersen, Rachel Christensen, Mike Coutts, Jennifer Byrd-Craven, Sara Ensenberger, Nicholas Geary, Larissa Haggard, Rebecca Hale, Mary Lemp, Patrick Maloney, Cy Nadler, Chattavee Numtee, Mahaley Ousley, Amanda Shocklee, and Ashley Stickney for help on various aspects of the project.

References

- Altman DG, Bland JM. Diagnostic tests 1: Sensitivity and specificity. *British Medical Journal* 1994;308:1552. [PubMed: 8019315]
- Barbarese WJ, Katusic SK, Colligan RC, Weaver AL, Jacobsen SJ. Math learning disorder: Incidence in a population-based birth cohort, 1976–82, Rochester, Minn. *Ambulatory Pediatrics* 2005;5:281–289. [PubMed: 16167851]
- Barrouillet P, Fayol M, Lathuli re E. Selecting between competitors in multiplication tasks: An explanation of the errors produced by adolescents with learning disabilities. *International Journal of Behavioral Development* 1997;21:253–275.
- Briars D, Siegler RS. A featural analysis of preschoolers' counting knowledge. *Developmental Psychology* 1984;20:607–618.
- Butterworth, B.; Reigosa, V. Information processing deficits in dyscalculia. In: Berch, DB.; Mazzocco, MMM., editors. *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities*. Baltimore: Paul H. Brookes Publishing Co; 2007. p. 65-81.
- Dehaene, S. *The number sense: How the mind creates mathematics*. New York: Oxford University Press; 1997.
- Dehaene S, Piazza M, Pinel P, Cohen L. Three parietal circuits for number processing. *Cognitive Neuropsychology* 2003;20:487–506.
- Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, et al. School readiness and later achievement. *Developmental Psychology* 2007;43:1428–1446. [PubMed: 18020822]
- Gallistel CR, Gelman R. Preverbal and verbal counting and computation. *Cognition* 1992;44:43–74. [PubMed: 1511586]

- Geary DC. Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin* 1993;114:345–362. [PubMed: 8416036]
- Geary DC. Mathematics and learning disabilities. *Journal of Learning Disabilities* 2004;37:4–15. [PubMed: 15493463]
- Geary, DC. Development of mathematical understanding. In: Kuhl, D.; Siegler, RS.; Damon, W., editors. *Cognition, perception, and language, Vol 2: Handbook of child psychology*. 6. New York: John Wiley & Sons; 2006. p. 777-810.
- Geary DC, Bow-Thomas CC, Liu F, Siegler RS. Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development* 1996;67:2022–2044. [PubMed: 9022227]
- Geary DC, Bow-Thomas CC, Yao Y. Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology* 1992;54:372–391. [PubMed: 1453139]
- Geary DC, Hamson CO, Hoard MK. Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology* 2000;77:236–263. [PubMed: 11023658]
- Geary DC, Hoard MK, Byrd-Craven J, Nugent L, Numtee C. Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development* 2007;78:1343–1359. [PubMed: 17650142]
- Gelman, R.; Gallistel, CR. *The child's understanding of number*. Cambridge, MA: Harvard University Press; 1978.
- Gersten R, Jordan NC, Flojo JR. Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities* 2005;38:293–304. [PubMed: 16122059]
- Gross-Tsur V, Manor O, Shalev RS. Developmental dyscalculia: Prevalence and demographic features. *Developmental Medicine and Child Neurology* 1996;38:25–33. [PubMed: 8606013]
- Jordan NC, Hanich LB, Kaplan D. Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology* 2003;85:103–119. [PubMed: 12799164]
- Koontz KL, Berch DB. Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition* 1996;2:1–23.
- Kosc L. Developmental dyscalculia. *Journal of Learning Disabilities* 1974;7:164–177.
- Landerl K, Bevan A, Butterworth B. Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition* 2003;93:99–125. [PubMed: 15147931]
- MacMillan, NA. Signal detection theory. In: Wixted, J.; Pashler, H., editors. *Stevens' handbook of experimental psychology*, 3rd Edition, Vol. 4: Methodology in experimental psychology. New York: John Wiley & Sons; 2002. p. 43-90.
- Murphy MM, Mazzocco MMM, Hanich LB, Early MC. 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* 2007;40:458–478. [PubMed: 17915500]
- National Mathematics Advisory Panel. *Foundations for success: Final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education; 2008. Available at <http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf>
- Ostad SA. Comorbidity between mathematics and spelling difficulties. *Logopedics Phoniatrics Vocology* 1998;23:145–154.
- Pickering, S.; Gathercole, S. *Working Memory Test Battery for Children (WMTB-C) manual*. London: The Psychological Corporation; 2001.
- Shalev RS, Manor O, Gross-Tsur V. Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine & Child Neurology* 2005;47:121–125. [PubMed: 15707235]
- Siegler RS. The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General* 1987;116:250–264.
- Siegler RS, Booth JL. Development of numerical estimation in young children. *Child Development* 2004;75:428–444. [PubMed: 15056197]
- Siegler RS, Ramani GB. Playing board games promotes low-income children's numerical development. *Developmental Science* 2008;11:655–661. [PubMed: 18801120]

- Siegler, RS.; Shrager, J. Strategy choice in addition and subtraction: How do children know what to do?. In: Sophian, C., editor. *Origins of cognitive skills*. Hillsdale, NJ: Lawrence Erlbaum; 1984. p. 229-293.
- Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCRC: Visualizing classifier performance in R. *Bioinformatics* 2005;21:3940–3941. [PubMed: 16096348]
- Walberg HJ. Improving the productivity of America's schools. *Educational Leadership* 1984;41:19–27.
- Wechsler, D. *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: PsychCorp, Harcourt Assessment; 1999.
- Wechsler, D. *Wechsler Individual Achievement Test-II*. San Antonio, TX: The Psychological Corporation, Harcourt Brace & Co; 2001a.
- Wechsler, D. *Wechsler Individual Achievement Test-II-Abbreviated*. San Antonio, TX: The Psychological Corporation, Harcourt Brace & Co; 2001b.
- Zorzi M, Priftis K, Umiltá C. Neglect disrupts the mental number line. *Nature* 2002;417:138. [PubMed: 12000950]

Circle all of the sets that add up to 5.
Work as quickly as you can.

5

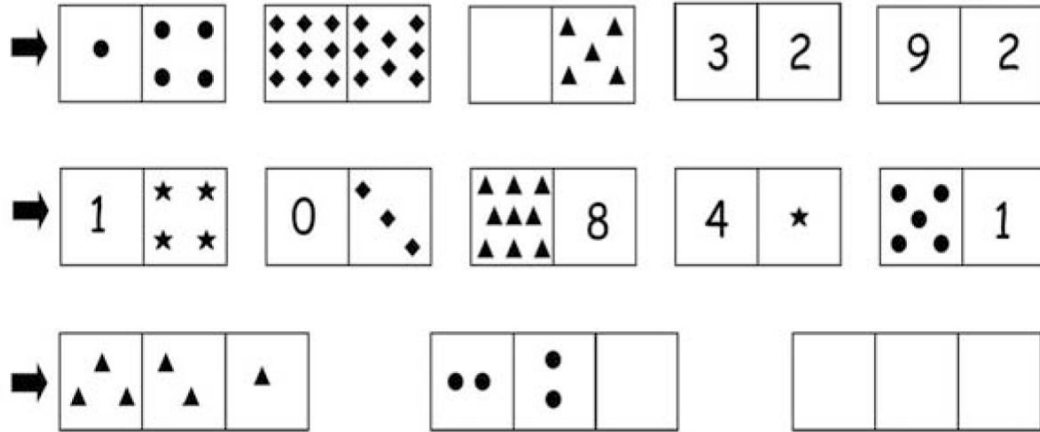


Figure 1.
Example Items From the Number Sets Test

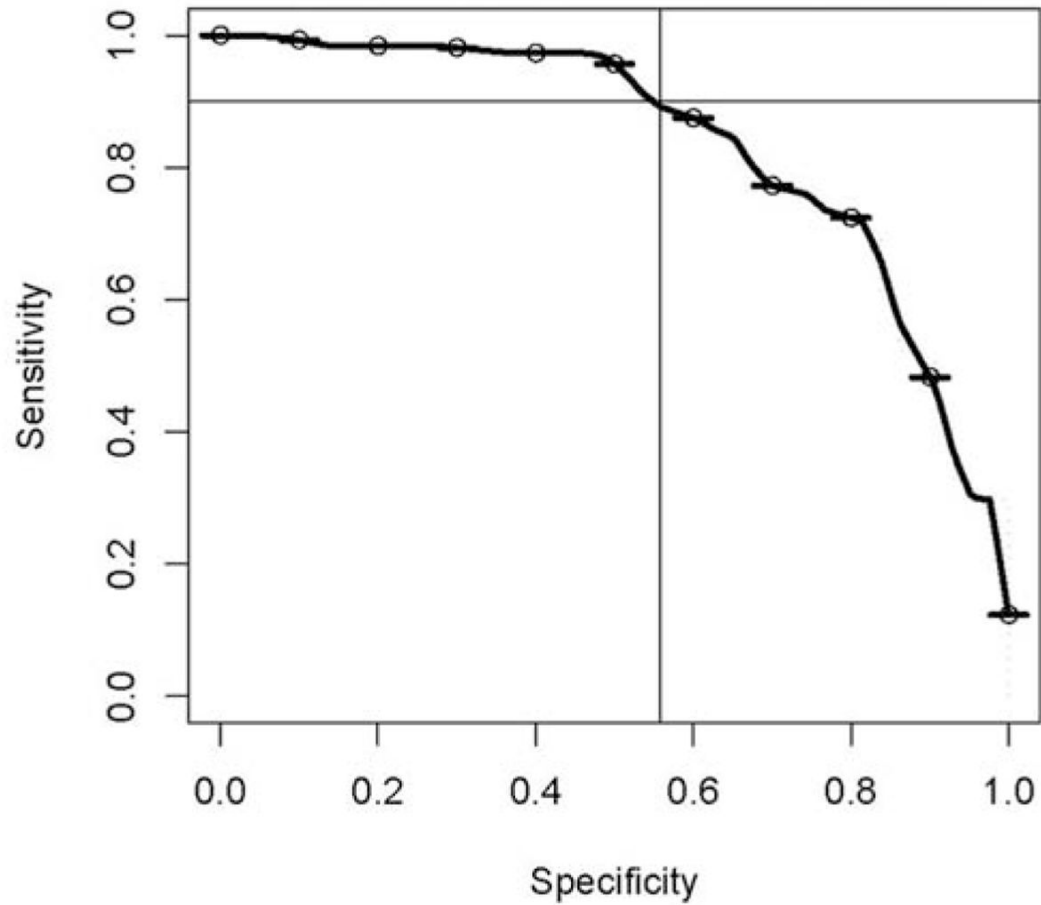


Figure 2.

The Response Operating Curve Using First Grade d' Scores From the Number Sets Test to Diagnosis MLD in Third Grade

Note: Sensitivity of d' for the correct classification of MLD is represented on the y-axis and specificity (incorrectly classifying a child as not having MLD) on the x-axis. The tradeoff between sensitivity and specificity is illustrated for a sensitivity of 0.9. Using d' scores to correctly identify 90% of children with MLD will result in 56% of children without MLD being correctly identified as such. MLD = mathematical learning disability.

Table 1Predictors of d' Scores

	Estimate	SE	<i>t</i>	<i>p</i>
Intercept	0.00	.068	< 1	< 1.00
IQ	0.27	.078	3.448	< .001
Central executive	0.34	.078	4.393	< .001
Correct retrieval	0.16	.075	2.155	.03
Counting error	0.23	.073	3.210	.002
Number line	-0.51	.083	-6.104	< .001

Table 2

Predictors of Third Grade Mathematics Achievement

	<i>Estimate</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	37.82	1.48	25.57	.001
First grade NO	4.73	1.73	2.74	.007
IQ	4.59	1.75	2.63	.009
Central executive	8.10	1.77	4.57	< .001
d'	5.01	1.99	2.52	.012

Note: NO = Number Operations national percentile ranking.

Table 3

Predictors of Third Grade Reading Achievement

	Estimate	SE	t	p
Intercept	56.48	0.90	62.83	< .001
First grade reading	20.49	1.15	17.76	< .001
IQ	0.67	1.11	< 1	> .50
Phonological Loop	3.09	1.09	2.82	.005
d'	-1.62	1.11	-1.46	> .10

Table 4

Mean Mathematics Achievement, IQ, and Cognition Performance

	Group		
	MLD	LA	TA
Mathematics achievement			
First grade	18	23	46
Second grade	8	22	63
Third grade	7	24	62
IQ	88	95	106
Cognition			
d'	-1.02	-0.41	0.50
Central Executive	-0.71	-0.30	0.44
Phonological Loop	-0.65	-0.25	0.36
Visuospatial Sketch Pad	-0.53	-0.19	0.30
Correct Retrieval	-0.47	-0.36	0.44
Counting Error	-0.70	-0.18	0.26
Number Line	0.85	0.45	-0.58

Note: Mathematics achievement scores are national (U.S.) percentile ranks, and IQ values are standardized scores from the test manual with $M = 100$, $SD = 15$. Cognition values are z scores with $M = 0$, $SD = 1$. High Number Line scores = inaccurate placement.

Table 5
 Categorization of Third Grade Achievement Status Using First Grade Mathematics Scores and d'

Diagnostic Measure	Predicted Group	Third Grade Group				Total
		MLD	L/A	NA		
First Grade Mathematics Achievement	MLD	18	1	4	23	
	L/A	19	11	30	60	
	NA	8	5	62	75	
d'	Hit	18	11	62	91	
	FA	5	49	13	67	
	CR	108	92	49	249	
	Miss	27	6	34	67	
	MLD	23	2	2	27	
d'	L/A	8	5	8	21	
	NA	14	10	86	110	
	Hit	23	5	86	114	
	FA	4	16	24	44	
	CR	109	125	38	272	
d'	Miss	22	12	10	44	

Note: MLD = mathematical learning disability; L/A = low achieving; NA = normally achieving; FA = false alarm; CR = correct rejection.

The bottom sections of Table 5 show the correspondence between the initial category based on the first grade d' scores and the diagnostic category at the end of third grade. In comparison to first grade achievement scores, the d' score correctly identified five additional children as MLD and overall produced more hits (114 vs. 91) and CRs (272 vs. 249), as well as less FAs (44 vs. 67) and misses (44 vs. 67).

Table 6Sensitivity and Specificity of First Grade Mathematics Achievement and d'

Diagnosing MLD	First Grade Achievement	d'
Initial cutoff		
Sensitivity	.40	.51
Specificity	.96	.96
Distribution cutoff		
Sensitivity	.69	.66
Specificity	.67	.88

Note: MLD = mathematical learning disability.