



Published in final edited form as:

J Learn Disabil. 2013 ; 46(5): 428–443. doi:10.1177/0022219411436214.

Prediction and Stability of Mathematics Skill and Difficulty

Rebecca B. Martin, MA¹, Paul T. Cirino, PhD¹, Marcia A. Barnes, PhD², Linda Ewing-Cobbs, PhD³, Lynn S. Fuchs, PhD⁴, Karla K. Stuebing, PhD¹, and Jack M. Fletcher, PhD¹

¹University of Houston, Houston, TX, USA

²University of Guelph, Guelph, Canada

³University of Texas at Houston Health Science Center, Houston, TX, USA

⁴Vanderbilt University, Nashville, TN, USA

Abstract

The present study evaluated the stability of math learning difficulties over a 2-year period and investigated several factors that might influence this stability (categorical vs. continuous change, liberal vs. conservative cut point, broad vs. specific math assessment); the prediction of math performance over time and by performance level was also evaluated. Participants were 144 students initially identified as having a math difficulty (MD) or no learning difficulty according to low achievement criteria in the spring of Grade 3 or Grade 4. Students were reassessed 2 years later. For both measure types, a similar proportion of students changed whether assessed categorically or continuously. However, categorical change was heavily dependent on distance from the cut point and so more common for MD, who started closer to the cut point; reliable change index change was more similar across groups. There were few differences with regard to severity level of MD on continuous metrics or in terms of prediction. Final math performance on a broad computation measure was predicted by behavioral inattention and working memory while considering initial performance; for a specific fluency measure, working memory was not uniquely related, and behavioral inattention more variably related to final performance, again while considering initial performance.

Keywords

math difficulty; reliable change; stability

The present study had two goals. First, we examined the stability of math performance over time, in terms of categorical (diagnostic) and continuous (reliable) dimensions, and how this stability is influenced by the severity of math difficulty and the measure used to define it. Second, we evaluated predictors of performance over time and the extent to which these predictors are differentially relevant in the context of the above factors. We begin by

Reprints and permissions: sagepub.com/journalsPermissions.nav

Corresponding Author: Rebecca B. Martin, University of Houston, 4800 Calhoun Road, Heyne Building, Houston, TX 77004, USA. Rmartin29@gmail.com.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

reviewing criteria for math difficulty (referred to as MD throughout this article), the rationale for subdividing MD into more versus less severe groups, methods for evaluating stability, measures used to establish math performance, and relevant predictors of math performance. Then, we present hypotheses.

Criteria for MD

Several methods have been employed for identifying learning disabilities. These include (a) IQ–achievement discrepancy, (b) performance below a percentile cut point on an achievement test (low achievement definitions or cutoff scores, including methods that require consistent low achievement; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007), (c) intraindividual differences (National Center for Learning Disabilities, 2002); and (d) response to intervention (RTI; Fletcher & Vaughn, 2009). The attention given to methods of identification is important given that the population studied as learning disabled and given access to interventions on the basis of those learning disabilities may vary according to the definition employed (Barbarese, Katusic, Colligan, Weaver, & Jacobsen, 2005). Despite the importance of accurate identification of learning disabilities, there is no current consensus regarding the “best” method. This is especially true in math, where less is known relative to reading in general and specifically with regard to identification, given a smaller math research base and a wide breadth of math skills. In this study, we establish MD on the basis of low achievement criteria because it plays a major role in many different identification procedures in clinical and research settings.

Severity

In the context of low achievement, there is no consensus regarding what constitutes an appropriate cut point, particularly given an underlying continuous distribution. Therefore, examining different levels of severity allows one to address the extent to which moving the cut point leads to quantitative or qualitative differences among groups on either side of the threshold, including the stability of their math performance. Several investigations have compared students who have very low math achievement scores (e.g., less than the 10th percentile) to those with performance within the low average range, finding more pervasive difficulties in the former group (Geary et al., 2007; Mazzocco & Kover, 2007; Murphy, Mazzocco, Hanich, & Early, 2007; Raghubar et al., 2009). Dichotomizing a continuous distribution may clarify differences between the groups that are formed, but it introduces statistical artifacts; for example, forming subgroups on the basis of correlated cognitive tasks complicates the interpretation of the subgroup with both cognitive difficulties having poorer performance on the outcome (e.g., Compton, DeFries, & Olson, 2001; Schatschneider, Carlson, Francis, Foorman, & Fletcher, 2002). However, individual difference characteristics important for math (e.g., working memory) can still potentially operate differentially in students with more versus less severe difficulties.

Stability 1: Categorical (Diagnostic) Change

Focusing on the issue of the stability of math performance (within or across cut points) is particularly relevant given the heterogeneity of math measures and what is assessed at

different developmental time points (Silver, Pennett, Black, Fair, & Balise, 1999). Previous work has focused on *categorical* change, that is, whether individuals stay in their original group (e.g., diagnostic concordance) over time. Several studies have demonstrated that students do not necessarily remain in their original group, with some students no longer identified as MD, whereas other students not initially identified may later be found to have difficulties (Geary & Brown, 1991; Silver et al., 1999). An epidemiological study of children from ages 5 to 19 (Barbarese et al., 2005) found the prevalence of MD to vary depending on the identification criteria utilized (from 6% for regression-based discrepancy methods to 14% for low achievement). Because categorical change is common at earlier ages (e.g., K–1), many MD studies focus on older elementary students (Chong & Siegel, 2008; Geary, 1993; Jordan, Hanich, & Kaplan, 2003). However, few studies follow students longitudinally, though more such studies have appeared recently (e.g., Jordan et al., 2003; Jordan, Kaplan, Locuniak, & Ramineni, 2007; Mazzocco & Thompson, 2005).

Stability 2: Continuous (Reliable) Change

In addition to categorical change, stability may also be indexed according to whether *continuous* scores change over time. For students with MD or learning difficulties in general, continuous change is most often expected to be positive in direction (even if *less* positive relative to students without MD); that is, in experimental studies, intervention students are expected to outperform their relevant controls. Few studies have as their primary goal to examine continuous change over time, and even fewer focus on students with difficulty. Test–retest reliability (and practice effect information) is certainly available for most broadband standardized instruments, but these data are primarily for a subsample of the norming group (often within the average range) and focused on group-level (rather than individual-level) change. Cirino et al. (2002) examined the stability of experimental and standardized measures in a sample of young students with reading difficulty and found few systematic practice effects and adequate test–retest reliability. Even here though, individual-level change was not evaluated, and analogous studies within the math domain are lacking. Continuous scores close to the population mean are less likely to change; more extreme (high or low) scores are expected to evidence more regression to the mean, which further complicates efforts to assess stability.

The reliable change index (RCI) is often used to assess the significance of continuous change at the individual level (Chelune, 2008; Jacobson & Truax, 1991; Jacobson, Truax, & Kazdin, 1992) and is judged relative to change in a relevant control group (to account for measurement error). In populations such as those with epilepsy or Alzheimer's disease, the clinical goal is to detect a meaningful *decline* in cognitive scores, though the technique was developed to detect reliable *increases* in mental health scores resulting from psychotherapy.

Three types of RCI are classic RCI (Jacobson & Truax, 1991), RCI correction (for practice effects, Chelune, Naugle, Luders, Sedlack, & Awad, 1993; for measurement error, Chelune, 2003), and RCI modified (RCIm), which considers different standard errors of measurement (SEM) for both an initial and a later time point (Iverson, 2001). Other variants have also been discussed (e.g., Marsden et al., 2011; McGlinchley, Atkins, & Jacobsen, 2002). Classic RCI uses the SEM from an initial time point to compute the standard error of the difference

(SE_{diff}), which describes the change scores' distribution if there were no actual change (Blasi et al., 2009). The SE_{diff} is then used to determine the confidence interval by multiplying it by 1.65 (since it is most commonly 90%). An RCI greater than 1.65 without real change would be expected to occur only 10% of the time; therefore, statistically rare changes can be defined as those that exceed this expected amount (Zabel, von Thomsen, Cole, Martin, & Mahone, 2009). Corrected RCI permits practice effects to be accounted for by adding the difference of the retest and baseline score (practice effect) to the retest score and adjusting the distribution around the new score. Last, RCIm suggests using the SEM from the initial time point (baseline) and the SEM from the later time point (retest) in computing the SE_{diff} instead of doubling the baseline SEM. This would account for any change in the standard deviation of the scores over time.

The present study was naturalistic (i.e., not an intervention study) and standard scores were used (as opposed to raw scores), and so we expected negligible practice effects and therefore did not require the "corrected" RCI. However, because different SEMs at test and retest are possible, we determined that the RCIm approach, which is well represented in the literature, had the best fit with our data (Iverson, Brooks, Collins, & Lovell, 2006; Martin et al., 2002; Ryan, Glass, Sullivan, Gibson, & Bartels, 2009; Zabel et al., 2009).

Continuous change can also be assessed with the standardized regression method (McSweeney, Naugle, Chelune, & Luders, 1993); regression analyses predict retest scores from baseline scores and then transform the residuals into standardized z scores by dividing by the standard error of the estimate (Blasi et al., 2009). The advantage is that the resulting z scores consider regression artifacts (regression to the mean) that are present whenever two measures with nonperfect correlations are evaluated (e.g., Campbell & Kenny, 1999). Typically, estimates of residualized change are derived from a control group and then applied to a clinical group, but this method is not applicable in the present context because definition and predictor variables were confounded. That is, math scores at the initial time point were used to define subgroups (designation as MD involves cut points on the continuous distribution) as well as to predict performance at a later time; therefore, the percentage showing "unusual" change is derived from a single population and so could not exceed the nominal rate. Some studies though have found strong agreement between various methods of defining meaningful change under various conditions (e.g., Atkins, Bedics, McGlinchey, & Beauchaine, 2005; Marsden et al., 2010).

Criterion Measures for Identification

For low achievement methods of defining MD, the key assessment tool is often a measure of computational skill (Barnes et al., 2002; Mabbott & Bisanz, 2008), particularly within the research literature. The use of such measures is not without problems. First, the use of common norm-referenced measures that assess many aspects of math may contribute to the heterogeneity of students identified with MD (Geary, 2004). For example, on measures such as the *Wide Range Achievement Test–4th Edition* (WRAT-4; Wilkinson & Robertson, 2006), *Wechsler Individual Achievement Test–3rd Edition* (Wechsler, 2009), *Woodcock–Johnson Tests of Academic Achievement–III* (WJ-III; Woodcock, McGrew, & Mather, 2001), and *Kaufman Test of Educational Achievement–2nd Edition* (Kaufman & Kaufman, 2004), math

computation subtests include items of arithmetic calculations with integers and real numbers as well as algebra, geometry, and calculus. Second, recent longitudinal research suggests that direct assessment of core deficits in MD, such as math fact retrieval (Geary & Hoard, 2001) or number sense (Gersten & Chard, 1999; Jordan, Kaplan, Nabors Olah, & Locuniak, 2006), likely are more important for identifying MD, particularly at younger ages. Finally, norm-referenced measures of computation skill may fail to capture individuals with difficulties in other areas of mathematics, for example, those with problem-solving weakness who may also have different patterns of cognitive functioning (Fuchs et al., 2008). Despite these difficulties, computational measures are most commonly used, and alternative measures are also not without problems. Therefore, we focus on a general, broadband measure of computation but also evaluate results with a measure of arithmetic fluency, which assesses specific skills that are more closely related to fact retrieval than broader based assessments.

Predictors: Demographic

Characteristics and Individual Differences

Demographic characteristics, such as socioeconomic status and sex, have been shown to predict the level of mathematical performance as well as rate of growth (Jordan et al., 2006; Jordan, Kaplan, Ramineni, & Locuniak, 2009). In addition, studies of students with MD with or without comorbid reading difficulties (Fuchs & Fuchs, 2002; Jordan et al., 2003; Powell, Fuchs, Fuchs, Cirino, & Fletcher, 2009; Silver et al., 1999) suggest larger differences on math word problems but more similar performance on arithmetic fact skills, implicating language in math development. Furthermore, reading and math skills are strongly correlated in typical students ($\sim r = .60$), which highlights the need to consider reading skills when studying math performance (Shapiro, Keller, Lutz, Santoro, & Hintze, 2006). These studies demonstrate the importance of assessing reading performance as both a categorical and a continuous variable in relation to math achievement.

In addition, cognitive skills such as attention and working memory have also been related to math skills. Cognitive attention assessed by continuous performance tests has been identified as a predictor of math calculation skill in children (Huckeba, Chapieski, Hiscock, & Glaze, 2008). On balance, though, more consistent relations with overall math competence have been shown with *behavioral* rating scales of inattentive behavior (Cirino, Fletcher, Ewing-Cobbs, Barnes, & Fuchs, 2007; Compton, Fuchs, Fuchs, & Bryant, 2006; Fuchs et al., 2005). Working memory also has been consistently linked to math performance (Bull, Espy, & Wiebe, 2008; De Smedt et al., 2009; Gathercole, Pickering, Knight, & Stegmann, 2004; Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Swanson & Kim, 2007). All three working memory components of the Baddeley and Hitch (1974) model (phonological loop, visuospatial sketchpad, and central executive) have been implicated. In many studies, the central executive, which has a supervisory role over the integration of information from the visuospatial sketchpad and phonological loop (Wu et al., 2008), plays a major role and has been shown to be related to skills such as number storage (Andersson & Lyxell, 2007) and concurrent counting tasks (Geary, Hoard, & Hamson, 1999; Hitch & McAuley, 1991) as well as overall achievement (Lehto, 1995).

Although much is known regarding factors that are predictive of math performance (reading, behavioral inattention, working memory), further replication is still needed, as many studies are selective in terms of the range of predictors used or examine only concurrent prediction (Alloway, 2009; Andersson & Lyxell, 2007; Fuchs et al., 2005; Geary et al., 1999; Jordan et al., 2007; Lehto, 1995; Siegel & Ryan, 1989). Furthermore, few studies have examined the extent to which such predictors may differ according to group designation or severity level.

Purpose and Hypotheses

A major goal of this study was to evaluate categorical (diagnostic) change and compare this to the RCI_m approach of continuous (reliable) change in math performance over 2 years. In doing so, we focused on students in later elementary school, after arithmetic skills are firmly established. In terms of categorical stability, we expected that children with MD would be more likely to change categorical group than would children without difficulty (no LD), given the more extreme and restricted range of scores of the former group. With regard to continuous stability, we expected that students with MD would be more likely to change positively on RCI_m (i.e., to significantly improve in math performance more than expected) over the 2-year period relative to no LD, but we did not expect differences in negative RCI_m change. Subdividing MD students into those with more versus less severity, we expected the more severe group to show *less* change in diagnostic category but a greater incidence of (positive) RCI change, relative to those in the LA group. We also evaluated the extent to which RCI changes would be minimized when estimated true scores were utilized. Within each type of change (categorical and continuous), we evaluated hypotheses with both a broad measure of calculation (which in this age range assesses primarily arithmetic but also procedural computations) and a specific measure of fluent arithmetic performance confined to single-digit calculations; however, we did not expect differences in terms of our stability hypotheses according to measure.

A second goal was to evaluate whether other demographic, academic, cognitive, and behavioral performance are predictive over time. We expected to replicate previous work that shows the relations of these characteristics and performances with math concurrently, as well as 2 years later, in the sample as a whole. We also expected that students' original categorization (as having difficulty and/or the severity of that difficulty) would not alter the predictive pattern; in other words, we did not expect an interaction between the predictors and group or subgroup.

Method

Participants and MD

Low Achievement Criteria—Participants were students in Grade 3 or Grade 4 from two urban school districts in two states (Tennessee and Texas) who were selected from a larger ($N = 291$) longitudinal study on math difficulties. In that larger study, students were tested at four time points; only the first (spring 2004) and last (in spring 2006) time points of data were utilized in this study and are referred to as the “initial” and “final” time points throughout this work, for ease of discussion, and as we were most interested in long-term stability. Students were also classified in the larger study as having math, reading, both math

and reading, or no learning difficulty. Because our focus was on change within mathematics over time, students with only reading difficulties ($n = 66$) and those not followed for the full 2 years (typically resulting from transferring from the school and thus missing data at the final time point; $n = 81$) were not included. The present study included the remaining 144 participants. Children who did not complete the study were compared to those who did on the characteristics in Table 2 (see below). Within those originally classified as MD or no LD, students were highly similar; however, no LD students who left the study had lower WRAT-3 Arithmetic scores. The overall proportion of students among those who left and those who remained was similar; however, for MD, those who left the study had higher free lunch rates, whereas among no LD, those who remained in the study had higher free lunch rates.

MD at the initial time point was characterized by math performance (WRAT-3 Arithmetic or WJ-III Math Fluency) below the 32nd percentile (standard score of 92) and WRAT-3 Reading subtest performance above the 40th percentile; controls performed above the 40th percentile on both math and reading measures (these criteria were used in the larger study, and the liberal cutoffs are consistent with other work; Cirino et al., 2007; Geary, Bow-Thomas, & Yao, 1992; Jordan et al., 2003). All children also had performance on one of the two subtests of the brief form of the *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) that was above the 2nd percentile ($T = 30$). Because we were interested in the influence of the measures used, we formed two subgroups with MD and two subgroups of controls, though with strong overlap (see Table 1). Of the 119 students classified by both measures, 84% of students were similarly identified as MD, $\chi^2(1, N = 119) = 54.92, p = .0001, \phi$ coefficient = .68.

Table 1 shows 83 students with MD (and 81 controls) according to the WRAT-3 Arithmetic subtest and 64 students (and 55 controls) according to the WJ-III Math Fluency subtest. Fewer students overall were classifiable under WJ-III criteria given the operational group criteria used in the larger study, which was based on the WRAT-3. Therefore, evaluation of categorical groups for this measure used this smaller sample, whereas evaluation of predictive influences for this measure relied on the total group as well as the reduced sample. In further analyses on MD, we also subdivided these groups into those with less severe difficulties (math performance between the 16th to 32nd percentiles) and those with more severe difficulties (below the 16th percentile; see Raghobar et al., 2009, for a similar grouping). Reading performance was not used to further subdivide groups but rather was included as a continuous predictor in the models presented below. Participant demographic data and performances under WRAT-3 Arithmetic criteria are presented in Table 2 (with similar results when comparisons were made according to WJ-III Math Fluency, except that those groups did not differ on working memory measures).

Measures—All measures were administered at the initial time point. For purposes of this study, the only measures utilized at the final time point were the math achievement measures.

The WRAT-3 (Wilkinson, 1993) is a frequently used measure of academic achievement assessing math, word reading, and spelling. We used two subtests. Arithmetic involves

number identification, counting, number comparisons, and other number tasks for young children; at school age, the task is primarily computations of increasing difficulty. Reading includes letter identification for young children; at school age, the task consists of word identification of increasing difficulty. Median internal consistency reliability of Arithmetic is .86; for Reading, .91. Test–retest reliability for Arithmetic is .87 (the value used to compute the RCI); for Reading, the test–retest reliability is .93.

WJ-III: The Math Fluency subtest of the WJ-III (Woodcock et al., 2001), a commonly used achievement battery, was utilized as our more specific test of basic arithmetic skill. For this subtest, participants solve as many single-digit problems (addition, subtraction, and multiplication) as possible in 3 min. Standard scores were used in analyses. Immediate test–retest reliability is .95 in a small sample of students, and 1-year test–retest reliability with a larger sample is reported to be .86 (McGrew & Woodcock, 2001); however, these latter scores were not used for reliable change as standard deviations were not provided. Standard errors of measurement are provided in the standard score metric and vary considerably (from 4.54 to 7.21) across our age range (McGrew & Woodcock, 2001); therefore, the standard errors corresponding to the students’ specific ages of testing were utilized for purposes of computing reliable change.

The Strengths and Weaknesses of ADHD and Normal Behavior (SWAN-IV; Swanson et al., 2005) is an 18-item teacher rating scale of inattention and hyperactivity or impulsivity rated on a 7-point Likert-type scale ranging from –3 to +3, with lower scores indicating more problematic behaviors. Each behavior corresponds to specific ADHD criteria identified in the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2000), with two factors: inattention (e.g., “gives close attention to detail and avoids careless mistakes”) and hyperactivity or impulsivity (e.g., “sits still, plays quietly”). We used the total score of the inattention scale, which has an alpha of .98 in the overall project sample, given its relation to mathematical outcomes (Cirino et al., 2007; Fuchs et al., 2005; Raghubar et al., 2010). One student was missing a score on this scale.

Working memory: We used the *Categorization Listening Span Task* and the *Visuospatial Working Memory Task* (Cornoldi, Marconi, & Vecchi, 2001) as measures of verbal and visual working memory, respectively. Categorization listening span requires the child to listen to strings of words, remember the last word in each group, tap whenever an animal name is presented, and repeat the final words in order. The stimuli include 16 trials. A single trial consists of presentation of a word string set of three or four words, with 4 trials at each of four levels (16 trials). Each trial of Level 1 requires recall of the last word of a single string; Level 2 requires recall of the last word of a set of two-word strings, in order; Levels 3 and 4 require recall of the last word of a set of three- or four-word strings, also in order. The Verbal Span Task has reduced linguistic demands in comparison to other verbal working memory tasks such as Listening Span (Daneman & Carpenter, 1980) because the stimuli are single words rather than complete sentences. The *Visuospatial Working Memory Task* is structured similarly, with four trials at each of four levels. Here, the experimenter touches three contiguous positions in a 4 × 4 matrix. The child taps the table if the positions are in a linear pattern (horizontal, vertical, or diagonal), remembering the last location touched in

each string showing the final locations on a blank matrix (recalling 1–4 locations). The verbal and visual working memory span tasks are designed to have similar response requirements but to differ only in the type of material to be remembered and manipulated. The structure of these tasks meets demand requirements similar to other often-utilized measures of working memory capacity, which have been shown to be useful in a variety of work (Conway et al., 2005). Examination of raw data indicated that students were in fact performing the secondary tasks (e.g., tapping when the word was an animal or the sequence was linear and refraining from doing so when it was not). The dependent measure from each task was a weighted total score depending on the number of words or locations to be recalled (e.g., Level 3 was weighted 3 because each trial required the recall of 3 words or locations to be considered correct).

Procedures—To assess *categorical* stability, students were reclassified at the final time point as MD or no LD, again using either WRAT-3 Arithmetic or WJ-III Math Fluency; a simple cut point was utilized (e.g., no LD were not required to have performances above the 40th percentiles in both reading and math as at the initial time point, only above the 32nd) to maximize sample size for comparisons. The MD groups were further subdivided into less severe (16th to 32nd percentile) and more severe (< 16th percentile).

To assess *continuous* stability, we calculated an RCIm (Iverson, 2001), which we illustrate for WRAT-3 Arithmetic. First, we calculated the SEM with the following formula, $SEM_1 = (1 - r_{12})^{1/2} * SD_1$, where r_{12} is the test–retest correlation coefficient of the normative sample (.87) and SD_1 is the standard deviation at the initial test–retest time point (15.3). Next, we computed the SEM at the final time point using the same formula with an r_{12} of .87 and a SD_2 of 14.7 (Wilkinson, 1993). Using these SEMs, we calculated the standard error of the difference ($SE_{diff} = [(SEM_1)^2 + (SEM_2)^2]^{1/2} = 7.65$). We then used the SE_{diff} to determine 90% confidence limits by multiplying it by 1.65, giving an RCI of 12.62. For WJ-III Math Fluency, we followed a similar procedure but used the standard score SEMs reported in the manual (McGrew & Woodcock, 2001) directly, according to the students' ages. The resultant 90% RCIs varied from 11.32 (for ages 11–13) to 14.90 (for ages 8–10). True-score estimates were computed by subtracting the population mean from the observed initial score and multiplying by reliability (.87 in the case of WRAT-3; .77 to .89 in the case of WJ-III Math Fluency) and then adding the population mean; for example, an observed score of 81 on the WRAT-3 yields a true-score estimate of 83.47.

Analyses—To evaluate our first hypothesis (change in performance), we defined groups *categorically* at the final time point according to the same criteria used at the initial time point, into MD and no LD using WRAT-3 Arithmetic. We then evaluated the extent to which the groups crossed the cut point over the time interval using the chi-square statistic. We further subdivided the students into more versus less severe subgroups and reevaluated categorical change among these subgroups. Then, we repeated these analyses, using WJ-III Math Fluency.

On the other hand, *continuous* change was noted when the difference in students' scores over time exceeded (increased or decreased relative to) the expected change (RCIm). Using this metric, we again compared the percentages of students with chi-square analyses, first

according to their original WRAT-3 Arithmetic classification as MD or no LD and then also compared subgroups of MD severity. These analyses were then repeated using groups classified on the basis of WJ-III Math Fluency.

To evaluate our second hypothesis, regression analyses were conducted to replicate relations previously found in the literature among working memory, behavioral inattention, and reading with math at both time points; initial math performance was also included in models at the final time point. Although the present sample was oversampled for math difficulties, the initial and final time point math performance scores were normally distributed, with the overall sample means only slightly below normative means. We also evaluated potential covariates of age, sex, ethnicity, and free lunch status; covariates were retained on the basis of their relationships with the math performance measures, predictor variables, and one another. We considered the extent to which WASI IQ influenced results, but substantive conclusions were unaltered. Then, variables were added to the model that designated level of math performance to determine whether relationships varied according to severity and for both criterion measures (WRAT-3 and WJ-III). Because we were most interested in how these factors influence the final outcome, these analyses were conducted only at the final time point.

Results

Categorical Change

Table 3 shows proportions of students who changed diagnostic groupings. For WRAT-3 Arithmetic, at the initial time point, 83 of the 144 students (58%) were classified as MD, but at the final time point, only 57 students (40%) were so classified. There was a significant difference in the proportion of students with MD versus no LD who changed, $\chi^2(1, N = 144) = 39.16, p < .0001$. The pattern was consistent with the hypothesis: Of the group with MD, 38.6% showed categorical change (i.e., they no longer met criteria), whereas only 9.8% of the no LD group showed categorical change (i.e., they now met criteria). Regarding severity, the three groups differed overall, $\chi^2(2, N = 144) = 17.84, p < .0001$; however, there was no significant difference in the proportion of students with more versus less severe initial difficulties who no longer met criteria for MD, $\chi^2(1, N = 83) = 2.39, p > .05$. The directional pattern of categorical change was, however, consistent with predictions (45% for less severe, 28% for more severe).

For WJ-III Math Fluency, at the initial time point, 64 of the 119 students (54%) met MD criteria, and at the final time point, 47 students (40%) met criteria. There was a significant difference in the proportion of students with categorical change, $\chi^2(1, N = 119) = 30.66, p < .0001$, and the pattern was similar to that just described: Students were more likely to move from MD to no LD (38%) than the reverse (13%). By severity group, overall categorical change was significant, $\chi^2(2, N = 119) = 15.64, p = .0004$; students with less severe initial difficulties were more likely to no longer meet criteria for MD (52%) relative to students with more severe difficulties (24%), $\chi^2(1, N = 64) = 5.11, p = .0238$.

Continuous Change

Table 4 shows the percentage of each group that changed in terms of RCI_m. For WRAT-3 Arithmetic, of the 144 students, 38 (26%) exhibited a continuous change. There was a significant difference between the MD and no LD groups, $\chi^2(2, N=144) = 16.14, p = .0003$; 34% of the no LD group showed continuous change compared to 20% of the MD group. For RCI_m, however, scores may either increase or decrease, so single degree of freedom contrast follow-up analyses were conducted to elucidate the effect. There were no group differences in the proportion who showed an improvement versus those who did not (those who remained stable or declined), $\chi^2(1, N=144) = 3.43, p > .05$; however, students in the no LD group were more likely to decline relative to the MD group (as opposed to remaining stable or improving), $\chi^2(1, N=144) = 14.45, p < .0001$. By severity group, overall continuous change was significant, $\chi^2(4, N=144) = 17.60, p = .0015$; however, among those with more versus less severe difficulties, there was no difference in the proportion who showed continuous change, $\chi^2(2, N=83) = 1.09, p > .05$; other comparisons were also not significant (all $p > .05$). When estimated true scores were used instead of initial observed scores, the overall proportion of students who exhibited any change was similar to the original analyses ($n = 34, 24\%$), although now the difference between the MD and no LD groups was not significant, $\chi^2(2, N=144) = 3.87, p > .05$. With observed scores, 29.51% of no LD and 6.02% of MD declined; here, though, the proportions were now 19.67% and 10.84%, respectively. By severity group, overall continuous change was also now not significant, $\chi^2(4, N=144) = 5.13, p > .05$.

When assessed with WJ-III Math Fluency, the pattern of continuous (RCI_m) change was not significant, $\chi^2(2, N=119) = 4.14, p > .05$. There were no differences in the proportion who improved, $\chi^2(1, N=119) = 2.25, p > .05$, declined, $\chi^2(1, N=119) = 2.48, p > .05$, or remained stable, $\chi^2(1, N=119) = 0.11, p > .05$. When evaluated by severity group, there were no overall differences in continuous change, $\chi^2(4, N=119) = 4.36, p > .05$; there were also no differences when only those with more versus less severe initial difficulties were compared, $\chi^2(2, N=64) = 0.34, p > .05$. The pattern of results did not change at all when estimated true scores were used instead of observed scores.

Predicting Math Performance

Characteristics from the initial assessment were evaluated as predictors of math performance at both time points. Table 5 shows the correlations between the predictors and outcomes. Evaluation of covariates revealed the following: Age was significantly related to both math performance measures at both time points, sex was related to WJ-III but not WRAT-3 scores (at both time points, with girls outperforming boys), free lunch status was related to WRAT-3 but not WJ-III scores (at both time points), and ethnicity was not related to math performance in this sample. Therefore, age and free lunch status were initially included in predictive models for WRAT-3 Arithmetic, and age and sex for WJ-III Math Fluency. However, 14 students had missing values for free lunch, and this variable did not contribute uniquely over age alone in predicting final WRAT-3 Arithmetic; therefore, this variable was not included in our models.

Table 6 shows regression results. At the initial time point, WRAT-3 scores were regressed on reading performance, verbal and visual working memory, and behavioral inattention. Consistent with our second hypothesis, these variables predicted 51% of the variability in WRAT-3 scores at the initial assessment, $F(5, 137) = 28.99, p < .0001$; reading performance ($p < .0001$) and inattention ($p < .0001$) were uniquely predictive (as was age), but working memory was not. Parameter estimates for predictors were positive, suggesting that better reading performance and less attention difficulty were associated with stronger math performance. The estimate for age was negative, however, with *lower* scores associated with *older* age. At the final time point, the autoregressor (initial WRAT-3 scores) was also included as a predictor; 55% of the variability was accounted for by the initial time point variables, $F(6, 136) = 27.30, p < .0001$; initial WRAT-3 performance, inattention, and both verbal and visual working memory performance (and age) demonstrated unique effects (all $p < .05$); reading did not.

Similar analyses were run with WJ-III Math Fluency (Table 6 shows values for the reduced sample—those who could be classified). At the initial time point in the full sample, the overall model was significant, $F(6, 136) = 19.54, p < .0001, R^2 = .46$. Reading ($p < .0004$), and inattention ($p < .0001$) as well as age ($p < .0006$) were uniquely predictive variables, with estimate interpretation similar to that for WRAT-3 Arithmetic. Results were highly similar in the reduced sample, $F(6, 111) = 22.87, p < .0001, R^2 = .55$, with a similar pattern of unique influences. For the final time point, the overall WJ-III Math Fluency model in the entire sample was also significant, $F(7, 135) = 23.78, p < .0001, R^2 = .55$; only the autoregressor ($p < .0001$) and inattention ($p < .04$) were uniquely predictive. The reduced sample was also significant, $F(7, 110) = 22.70, p < .0001, R^2 = .59$, but attention was no longer significant ($t = 1.72, p < .09$). The covariates were not uniquely contributory to either of the final time point models, and results were similar with or without them. We then also ran similar analyses that predicted difference scores over the time points, and although there are known issues with the use of such scores (Nunnally & Bernstein, 1993), results were substantively similar to those presented above.

To evaluate whether predictors operated differently within group, we repeated the analyses above predicting final time point performance but added interaction variables that crossed each predictor with initial performance. This is a proxy for whether students had MD and/or whether they were more or less severe. However, we also ran models that included the actual two- or three-group designation term in the models (MD vs. no LD; more severe MD vs. less severe MD vs. no LD), and the results were highly substantively similar; therefore, only the continuous term interactions are reported below. For WRAT-3 Arithmetic, the model including the interaction terms of predictor by initial performance resulted in a significant overall model, $F(11, 131) = 15.57, p < .0001, R^2 = .57$. However, none of the individual interaction terms was significant; the R^2 was .02 for all the interaction terms together and was not significantly larger than the model reported above without these terms ($p > .05$). For the WJ-III, only the reduced sample models could be evaluated since not all students were classifiable. The overall model with the interaction terms was significant, $F(13, 104) = 11.83, p < .0001, R^2 = .60$; however, none of the individual interaction terms was significant. The overall R^2 at .01 was also not significant ($p > .05$).

Discussion

The first goal of this study was to determine which factors, including categorical versus continuous change, liberal versus conservative cut point, and broad computational versus specific math fluency assessment, affect stability versus change in math performance in elementary school-children. The second goal was to evaluate the influence of other skills (reading, working memory, and attention) on concurrent and future math performance and to examine whether such predictors operated differentially among students with different initial designations as MD or no LD. The key contributions of the present study are the differentiation of categorical versus continuous change and its implications and the evaluation of predictive factors relevant for math that also considers how they might vary according to diagnostic group.

Categorical and Continuous Change

Generally, change depended on all three factors. Change was more common for categorical distinctions, for students closer to a set cut point relative to those farther from it, and for broader rather than more specific assessments. The question of how stable MD is over time has been previously addressed in other contexts (Badian, 1999; Geary & Brown, 1991; Mazzocco & Kover, 2007; Silver et al., 1999). In those studies, the focus has been on whether students change (categorically) rather than the degree of individual change, and multiple definitions are not often considered. In the present study, nearly 40% of the children who were initially identified as MD were recategorized to no LD over time, relative to about 10% of initial no LD students, who later met criteria for MD. This was true whether we assessed a broad (WRAT-3 Arithmetic) or a specific (WJ-III Math Fluency) computational measure. In a longitudinal study from Grade 1 to Grade 8, Badian (1999) classified 5.7% of the sample as having persistently low math achievement (mean standard score on a measure of math computation over a 7- or 8-year period that was < 25th percentile); in that study, there were lower prevalence rates of persistent difficulty in math only (3.9%) than for reading only (6.0%). However, defining *persistent* by mean scores makes it difficult to examine fluctuations from any given point or to account for scores close to the cut point. Silver et al. (1999) found that between one-third and one-half of the original group of children with arithmetic difficulties remained stable over time regardless of whether the size of IQ-achievement discrepancy was 1 versus 1.5 standard deviations. Relative to Silver et al., we found higher rates of categorical stability overall (74%) as well as among those originally identified with difficulty (61%), though using a different identification procedure (low achievement) and a somewhat longer time frame (24 relative to 19 months).

However, categorical change can be misleading because it does not consider the extent to which scores fluctuate naturally given measurement error. The present study explicitly considered this type of continuous change by employing an RCI statistic, which to our knowledge has not previously been evaluated in the MD literature. Here, the proportion of students who changed continuously differed according to the type of measure examined and also differed from the rates of categorical change across subgroups. When a broad computational measure was utilized, overall continuous change was the same as categorical change (26%). However, within the MD subgroup, categorical change was twice as common

as continuous change (39% relative to 20%); students with MD were more likely than those with no LD to change (to increase past the criterion threshold). By contrast, students with no LD were more likely to show continuous change (to have scores decrease), though even here, this is likely the result of their greater distance from the cut point, which was in part accounted for by the true-score adjustment. When a more specific fluency measure was used, overall continuous change was similar to categorical change (27% relative to 26%), and this was generally true within the subgroup with MD specifically as well (28% compared to 38%). Similar to the results with the broad measure, students with MD were more likely to change than those with no LD categorically. Change was even more similar for the specific measure in these two subgroups when it was evaluated with RCI.

Our results concerning the difference in categorical versus continuous change may have implications for assessing responsiveness in RTI. For example, some studies refer to or specify a criterion to be considered “no longer at risk” for learning disability following intervention (Scanlon, Gelzheiser, Vellutino, Schatschneider, & Sweeney, 2008). However, there is not clear agreement regarding the correct identification of nonresponders versus responders (Barth et al., 2008). At the individual level, therefore, RCI may provide an alternative strategy for gauging responsiveness by determining whether the amount of change is reliable and potentially meaningful, independent of reaching the criterion goal. In some ways, this use is conceptually similar to the use of slope measures to gauge responsiveness, often used in combination with performance level (McMaster, Fuchs, Fuchs, & Compton, 2005; Speece & Case, 2001). Such a process becomes particularly important for struggling students since passing a static benchmark would become more difficult as the initial distance from the criterion grows and may supplement group-based differences in intervention studies. Clearly, however, in terms of external validity of a student’s RTI, benchmarks will likely remain important goals.

There were relatively few differences between students with less severe versus more severe MD. Across change metrics and types of criterion measures, the only significant finding was for categorical change on WJ-III Math Fluency, where students in the less severe MD group were more likely to change (52%) relative to those with more severe MD (24%); a similar though not significant pattern was noted for WRAT-3 Arithmetic (45% to 28%). However, given that there was no difference in the proportion of students who showed reliable (continuous) improvement, using either the broad or specific measure, the seemingly greater stability seen in the more severe MD group is likely more a function of their initial distance from a benchmark rather than fundamental rates of change versus stability.

Prediction of Performance Over Time

At the initial time point, overall models were strongly predictive ($R^2 = 51\%$ and 46%) of both broad computational and specific fluency skill; reading and inattention were uniquely predictive, but working memory was not, for both models. At the later time point, overall models were again strongly predictive ($R^2 = 54\%$ and 55%), although a different pattern emerged according to type of measure. For broad computations, initial score, inattention, and both verbal and visual working memory were all unique predictors. However, for the

more specific skill of computational fluency, only initial score was a unique predictor in the reduced models of Table 6.

Previous work has noted the utility of reading for prediction of broad computational or fluency measures (Fletcher, 2005). Geary et al. (2007) also found more severe word reading difficulties in their MLD group in kindergarten and first grade than in their low achieving and typically achieving math groups. The relation between word reading deficits and math performance has been explained through a possible deficit in processing speech sounds, which would affect both skills (Geary & Hoard, 2001; Jordan et al., 2003). The influence of reading, which we found on both math measures, was restricted to the initial time point, however. This is in contrast to Jordan et al. (2002), whose growth curve study indicated that reading skill was more likely to influence math growth than vice versa. However, the outcome in that case was a composite measure that included problem solving, which is more closely related to reading skills than are measures of computation or fluency.

The current study also highlights the role of inattention with regard to our mathematical outcomes, particularly for the broader computational measure. The impact of inattention predicting math performance has been found previously in several studies (Fuchs et al., 2006; Huckleba et al., 2008; Raghobar et al., 2009). However, the mechanism remains unclear. For example, it has been suggested that inattention contributes to certain types of mathematical errors, including switching from one operation to another on mixed format computation tasks (Raghobar et al., 2009), which can negatively affect math performance, although that study did not find inattention to relate directly to switch errors. There is also a distinction between the behavioral inattention assessed in both Raghobar et al. (2009) as well as the present study versus measures of cognitive attention. For example, measures of sustained attention (e.g., omissions on continuous performance tests) have been shown to be related to weaker computation performance (Lindsay, Tomazic, Levine, & Accardo, 2001). Clearly, more work is needed to understand specifically why and what type of inattention is predictive of math performance, which may also differ according to the measure evaluated. The present study found that inattention was not uniquely related to fluency skills at the final time point, though the zero-order relations of inattention with both types of math skill were similar at both time points, and in the full sample inattention was uniquely predictive.

Working memory has been found to be a predictor of math performance (Bull & Scerif, 2001) and to differentiate level of math performance (Gathercole, Pickering, Ambridge, & Wearing, 2004). Here, working memory was a unique predictor for broad math computations but not for specific math fluency. Although working memory is likely a strong contributor to the development of math fact learning and retrieval, at the grades studied here (3–6) its contribution may be more muted to the extent that such facts are established. An intriguing finding was that for broad math computations, working memory predicted later but not initial (concurrent) performance, whereas the opposite pattern was true for the reading predictor. At the initial time point, the majority of the types of problems solved are similar to those of the more basic fluency measures, where similar predictive patterns were evidenced. In contrast, more complex computation (such as those that students encountered at the final time point) requires a wider range of procedures and algorithms that require the storage,

processing, and manipulation of the contents of working memory to produce a correct answer.

Although predictors differed according to the *type* of math measure examined, we did not find a similar distinction among students with different *levels* of math performance (utilizing either continuous predictors or MD group designations). We did not find evidence that the predictors operated on final performance differentially according to initial performance. We therefore conclude that the predictors of math performance are similar across the continuum of skill level. These results are on the surface at odds with other studies that have found differences among students with more versus less severe MD (Geary et al., 2007; Geary et al., 2009; Murphy et al., 2007), where the implication is that cognitive skills operate differently in students with more severe MD (typically referred to as MLD in those studies) than in students with less severe MD (referred to as low achievement). However, differences in predictors employed and identification procedures can help account for differences in interpretation. For example, Geary et al. (2007) found that the central executive fully mediated the relationship of both counting knowledge and retrieval errors on complex addition with group severity. Those authors also reported that although working memory is predictive of math performance in MLD (< 15th percentile in both kindergarten and Grade 1) and LA groups (23rd to 39th percentiles), only the MLD group exhibited working memory deficits. In a later study, Geary et al. (2009) found that although number sense measures are significantly predictive of both MLD or LA membership, below-average IQ and working memory deficits explained the distinction between them.

It is also possible that the cut score we employed for the more severe MD group was still too liberal to find significant group differences. For example, Murphy et al. (2007) used a more stringent cutoff of less than 10% and found their MLD group to have a significantly lower rate of growth from kindergarten to Grade 3 than both their LA group and their typically achieving comparison group. Another example is seen in children with developmental or acquired dyscalculia, (cutoff < 5%) who have also been found to remain stable (Auerbach, Gross-Tsur, Manor, & Shalev, 2008; Temple & Sherwood, 2002). The latter study concluded that mathematical deficits are more related to number-based factors than cognitive ones such as working memory (Temple & Sherwood, 2002). These studies identify a small proportion (~5%) of students with MLD, primarily by requiring scores to be below a relatively stringent cutoff level at more than one time point. The results of the present study suggest that any group of students identified by such criteria may be more stable (when scores pass the categorical threshold but remain within the expected RCI) or less stable (when scores do not pass the categorical threshold but still show greater than expected RCI) than might be initially assumed, and this stability can be influenced by both where the threshold is set and what type of math measure is used to define groups.

Limitations and Future Directions

The present study expands the literature in a number of ways by considering multiple factors that influence stability of math scores among students with and without MD. Readers should, however, interpret findings with some important limitations in mind. These include the use of a liberal cutoff score, reliance on only computational measures to define the MD

group, inclusion of a relatively limited sample of predictors, and restriction to initial and final time points.

A liberal cutoff has been used and supported in several studies (Hanich et al., 2001; Geary et al., 2004; Jordan, Kaplan, & Hanich, 2002). Moreover, subgrouping of participants within this larger sample allowed for comparison of severity levels directly. Although using a composite measure or both (or additional) types of criterion math skills simultaneously may have advantages, the use of computational measures is still the most common way math difficulties are identified in research studies (Berg, 2008; Mabbott & Bisanz, 2008; Rosselli, Matute, Pinto, & Ardila, 2006; Siegel & Ryan, 1989; H. L. Swanson & Beebe-Frankenberger, 2004). Still, patterns of change and correlates may differ for problem solving and computations (Fuchs et al., 2008). Additional predictors could also have been utilized, particularly those focused on numbers that may be more specific in identifying students most likely to have the lowest scores because of an underlying lack of both procedural and conceptual number knowledge. Yet although instruments of this type have been more commonly used at younger ages (Cirino, 2011; Jordan et al., 2006; Krajewski & Schneider, 2009; LeFevre et al., 2006), they are less commonly available for the age of the students in this study. Nevertheless, this represents an interesting avenue for future research. Finally, given that the students in this study were evaluated on up to four occasions, it would be possible to track their performance with growth curves, as has been done in other longitudinal math research (Jordan et al., 2009). Although such studies are clearly informative and would no doubt offer another perspective on the issues considered here, the present study emphasized change from one point to another because this is the case when making a diagnosis or determining whether improvement has occurred. In sum, it would be beneficial to evaluate the hypotheses of the present study across multiple time points and with several outcomes, with a broader range of predictor variables, and using different definitions of MD.

A strength of this study is its consideration of an RCI for evaluating individual change and comparing this to categorical metrics. However, components of RCI were derived from normative manuals (McGrew & Woodcock, 2001; Wilkinson, 1993) and differed across measures. For broad computations, the criterion for change was established with a test–retest coefficient (and the standard deviations of that sample), whereas age-specific SEMs were utilized for the fluency measure. Ideally, comparisons across measures would compute the RCI in the same way, as would be the case in an intervention study with a randomized control group. Nonetheless, both measures are highly reliable, and the changes required to exceed RCI were similar (approximately 11–14 standard score points).

Conclusion

The present study highlights the impact cut-point designation and the choice of measure have on who is potentially identified as having difficulty. The findings of this study also suggest promise for the use of a continuous change measure of stability (RCI) over a categorical one (cut point). These continuous stability results were more similar than different across measures and for students with differing initial levels of performance; future research should evaluate other discrete measures in analogous manner to determine their

relevance both for identification and for establishing change. Further consideration of the emerging implications of MD as a categorical versus continuous dimension is also warranted. Taken together, such information may help to identify patterns of academic and cognitive skills that are more or less likely to remain stable over time and may point to productive avenues for designing interventions.

Acknowledgments

Funding

This research was supported in part by grant P01 HD04621 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or the National Institutes of Health.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed., text revision. Author; Washington, DC: 2000.
- Andersson U, Lyxell B. Working memory deficit in children with mathematical difficulties: A general or specific deficit? *Journal of Experimental Child Psychology*. 2007; 96:197–228. doi:10.1016/j.jecp.2006.10.001. [PubMed: 17118398]
- Alloway TP. Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment*. 2009; 25(2):92–98.
- Atkins DC, Bedics JD, McGlinchey JB, Beauchaine TP. Assessing clinical significance: Does it matter which method we use? *Journal of Consulting and Clinical Psychology*. 2005; 73(5):982–989. doi: 10.1037/0022-006X.73.5.982. [PubMed: 16287398]
- Auerbach JG, Gross-Tsur V, Manor O, Shalev RS. Emotional and behavioral characteristics over a six-year period in youths with persistent and nonpersistent dyscalculia. *Journal of Learning Disabilities*. 2008; 41:263–273. doi:10.1177/0022219408315637. [PubMed: 18434292]
- Baddeley, A.; Hitch, GJ. Working memory. In: Bower, GA., editor. *Recent advances in learning and motivation*. Academic Press; New York, NY: 1974. p. 47-90.
- Badian NA. Persistent arithmetic, reading, or arithmetic and reading disability. *Annals of Dyslexia*. 1999; 49:45–70.
- 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. doi:10.1367/A04–209R.1. [PubMed: 16167851]
- Barnes MA, Pengelly S, Dennis M, Wilkinson M, Rogers T, Faulkner H. Mathematics skills in good readers with hydrocephalus. *Journal of the International Neuropsychological Society*. 2002; 8:72–82. [PubMed: 11843076]
- Barth A, Stuebing KK, Anthony J, Denton C, Mathes P, Fletcher JM, Francis D. Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences*. 2008; 18:296–307. doi:10.1016/j.lindif.2008.04.004. [PubMed: 19081758]
- Berg DH. Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology*. 2008; 99:288–308. doi:10.1016/j.jecp.2007.12.002. [PubMed: 18242631]
- Blasi S, Zehnder AE, Berres M, Taylor KI, Spiegel R, Monsch AU. Norms for change in episodic memory as a prerequisite for the diagnosis of mild cognitive impairment (MCI). *Neuropsychology*. 2009; 23:189–200. [PubMed: 19254092]
- Bull R, Espy KA, Wiebe SA. Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*. 2008; 33:205–228. doi:10.1080/87565640801982312. [PubMed: 18473197]

- Bull R, Scerif G. Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*. 2001; 19:273–293. [PubMed: 11758669]
- Campbell, DT.; Kenny, DA. A primer on regression artifacts. Guilford Press; New York, NY US: 1999.
- Chelune, GJ. Assessing reliable neuropsychological change.. In: Franklin, RD., editor. Prediction in forensic and neuropsychology: Sound statistical practices. Lawrence Erlbaum; Mahwah, NJ: 2003. p. 123-147.
- Chelune GJ. Evidence-based research and practice in clinical neuropsychology. *Clinical Neuropsychology*. 2008; 24(3):1–14. doi:10.1080/13854040802360574.
- Chelune GJ, Naugle RI, Luders H, Sedlack J, Awad IA. Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*. 1993; 7:41–52.
- Chong SL, Siegel LS. Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology*. 2008; 33:300–317. [PubMed: 18473201]
- Cirino PT. The interrelationships of mathematical precursors in kindergarten. *Journal of Experimental Child Psychology*. 2011; 108(4):713–733. [PubMed: 21194711]
- Cirino PT, Fletcher JM, Ewing-Cobbs L, Barnes MA, Fuchs LS. Cognitive arithmetic differences in learning difficulty groups and the role of behavioral inattention. *Learning Disabilities Research & Practice*. 2007; 22:25–35.
- Cirino PT, Rashid FL, Sevcik RA, Lovett MW, Frijters JC, Wolf M, Morris RD. Psychometric stability of nationally normed and experimental decoding and related measures in children with reading disability. *Journal of Learning Disabilities*. 2002; 35:525–538. [PubMed: 15493250]
- Compton DL, DeFries JC, Olson RK. Are RAN-and phonological awareness-deficits additive in children with reading disabilities? *Dyslexia*. 2001; 7:125–149. [PubMed: 11765981]
- Compton DL, Fuchs D, Fuchs LS, Bryant JD. Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*. 2006; 98:394–409.
- Conway ARA, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW. Working memory span tasks: A methodological review and user's guide. *Psycho-nomic Bulletin & Review*. 2005; 12:769–786.
- Cornoldi C, Marconi F, Vecchi T. Visuospatial working memory in Turner's system. *Brain and Cognition*. 2001; 46(Tennet 11):90–94. [PubMed: 11527371]
- Daneman M, Carpenter PA. Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*. 1980; 19:450–466.
- De Smedt B, Janssen R, Bouwens K, Verschaffel L, Boets B, Ghesquière P. Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology*. 2009; 103:186–201. [PubMed: 19281998]
- Fletcher JM. Predicting math outcomes: Reading predictors and comorbidity. *Journal of Learning Disabilities*. 2005; 38:308–312. [PubMed: 16122061]
- Fletcher JM, Vaughn S. Response to intervention models as alternatives to traditional views of learning disabilities: Response to the commentaries. *Child Development Perspectives*. 2009; 3:48–50.
- Fuchs LS, Compton DL, Fuchs D, Paulsen K, Bryant JD, Hamlett CL. The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*. 2005; 97:493–513.
- Fuchs LS, Fuchs D. Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities*. 2002; 35:563–573. [PubMed: 15493253]
- Fuchs LS, Fuchs D, Compton DL, Powell, Seethaler PM, Capizzi AM, Schatschneider C, Fletcher JM. The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*. 2006; 98:29–43.
- Fuchs LS, Fuchs D, Stuebing KK, Fletcher JM, Hamlett CL, Lambert W. Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology*. 2008; 100:30–47. [PubMed: 20057912]
- Gathercole SE, Pickering SJ, Ambridge B, Wearing H. The structure of working memory from 4 to 15 years of age. *Developmental Psychology*. 2004; 40:177–190. [PubMed: 14979759]

- Gathercole SE, Pickering SJ, Knight C, Stegmann Z. Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*. 2004; 18:1–16.
- Geary DC. Mathematical disabilities: Cognitive, neuro-psychological, 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, Baily DH, Littlefield A, Wood P, Hoard MK, Nugent L. First-grade predictors of mathematical learning disability: A latent class trajectory analysis. *Cognitive Development*. 2009; 24:411–429.
- 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, Brown SC. Cognitive addition: A short-term longitudinal study of strategy choice and speed-of-processing. *Developmental Psychology*. 1991; 27:787.
- Geary DC, Hoard MK. Numerical and arithmetical deficits in learning-disabled children: Relation to dyscalculia and dyslexia. *Aphasiology*. 2001; 15:635–647.
- Geary DC, Hoard MK, Byrd-Craven J, DeSoto MC. Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*. 2004; 88:121–151. doi:10.1016/j.jecp.2004.03.002S0022096504000335. [PubMed: 15157755]
- 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. doi:10.1111/j.1467-8624.2007.01069.x. [PubMed: 17650142]
- Geary DC, Hoard MK, Hamson CO. Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for mathematical disability. *Journal of Experimental Child Psychology*. 1999; 74:213–239. [PubMed: 10527555]
- Gersten R, Chard D. Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education*. 1999; 33:18.
- Hanich LB, Jordan NC, Kaplan D, Dick J. Performance across different areas of mathematical cognition in children with Learning Difficulties. *Journal of Educational Psychology*. 2001; 93(3): 615–626.
- Hitch GJ, McAuley E. Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology*. 1991; 82:375–386. [PubMed: 1954527]
- Huckeba W, Chapieski L, Hiscock M, Glaze D. Arithmetic performance in children with Tourette syndrome: Relative contribution of cognitive and attentional factors. *Journal of Clinical and Experimental Neuropsychology*. 2008; 30:410–420. [PubMed: 18938679]
- Iverson GL. Interpreting change on the WAIS-III/WMSIII in clinical samples. *Archives of Clinical Neuropsychology*. 2001; 16:183–191. [PubMed: 14590186]
- Iverson GL, Brooks BL, Collins MW, Lovell MR. Tracking neuropsychological recovery following concussion in sport. *Brain Injury*. 2006; 20:245–252. [PubMed: 16537266]
- Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*. 1991; 59:12–19. [PubMed: 2002127]
- Jacobson, NS.; Truax, P.; Kazdin, AE. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research.. In: Kazdin, AE., editor. *Methodological issues & strategies in clinical research*. American Psychological Association; Washington, DC: 1992. p. 631–648.
- Jordan NC, Hanich LB, Kaplan D. A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development*. 2003; 74:834–850. [PubMed: 12795393]
- Jordan NC, Kaplan D, Hanich LB. Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology*. 2002; 92(3):586–598.

- Jordan NC, Kaplan D, Locuniak MN, Ramineni C. Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice*. 2007; 22:36–46.
- Jordan NC, Kaplan D, Nabors Olah L, Locuniak MN. Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*. 2006; 77:153–175. doi:10.1111/j.1467–8624.2006.00862.x. [PubMed: 16460531]
- Jordan NC, Kaplan D, Ramineni C, Locuniak MN. Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*. 2009; 45:850–867. doi: 10.1037/a0014939. [PubMed: 19413436]
- Kaufman, AS.; Kaufman, NL. Kaufman Test of Educational Achievement administration manual. 2nd ed.. Psychological Corporation; San Antonio, TX: 2004.
- LeFevre JA, Smith-Chant BL, Fast L, Skwarchuk SL, Sargla E, Arnup JS, Penner-Wilger M, Bisanz J, Kamawar D. What counts as knowing? The development of conceptual and procedural knowledge of counting from kindergarten through Grade 2. *Journal of Experimental Child Psychology*. 2006; 93(4):285–303. [PubMed: 16360166]
- Lehto J. Working memory and school achievement in the ninth form. *Educational Psychology*. 1995; 15:271–283.
- Lindsay RL, Tomazic T, Levine MD, Accardo PJ. Attentional function as measured by a Continuous Performance Task in children with dyscalculia. *Journal of Developmental and Behavioral Pediatrics*. 2001; 22:287–292. [PubMed: 11718231]
- Mabbott DJ, Bisanz J. Computational skills, working memory, and conceptual knowledge in older children with mathematics learning disabilities. *Journal of Learning Disabilities*. 2008; 41:15–28. [PubMed: 18274501]
- Marsden J, Eastwood B, Wright C, Bradbury C, Knight J, Hammond P. *Addiction*. 2011; 106(2):294–302. doi: 10.1111/j.1360-0443.2010.03143. [PubMed: 20925686]
- Martin R, Sawrie S, Gilliam F, Mackey M, Faught E, Knowlton R, Kuzniecky R. Determining reliable cognitive change after epilepsy surgery: Development of reliable change indices and standardized regression-based change norms for the WMS-III and WAIS-III. *Epilepsia*. 2002; 43(12):1551–1558. [PubMed: 12460258]
- Mazzocco MM, Kover ST. A longitudinal assessment of executive function skills and their association with math performance. *Child Neuropsychology*. 2007; 13:18–45. [PubMed: 17364562]
- Mazzocco MM, Thompson RE. Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice*. 2005; 20:142–155. [PubMed: 20084182]
- McGlinchey JB, Atkins DC, Jacobson NS. Clinical significance methods: Which one to use and how useful are they? *Behavior Therapy*. 2002; 33:529–550.
- McGrew, KS.; Woodcock, RW. Technical manual: Woodcock–Johnson III. Riverside; Itasca, IL: 2001.
- McMaster KL, Fuchs D, Fuchs LS, Compton DL. Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*. 2005; 71:445–463.
- McSweeney AJ, Naugle RI, Chelune GJ, Luders H. “T scores for change”: An illustration of a regression approach to depicting change in clinical neuropsychology. *Clinical Neuropsychologist*. 1993; 7:300–312.
- Murphy MM, Mazzocco MM, 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 Center for Learning Disabilities. *Achieving better outcomes—maintaining rights: An approach to identifying and serving students with specific learning disabilities*. Author; New York, NY: 2002.
- Nunnally, JC.; Bernstein, IH. *Psychometric theory*. 3rd Ed.. McGraw-Hill; New York: 1993.
- Powell SR, Fuchs LS, Fuchs D, Cirino P, Fletcher J. Effects of fact retrieval tutoring on third-grade students with math difficulties with and without reading difficulties. *Learning Disabilities Research & Practice*. 2009; 24:1–11. [PubMed: 19448840]
- Raghubar KP, Barnes MA, Hecht SA. Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning & Individual Differences*. 2010; 20(2): 110–122.

- Raghubar KP, Cirino P, Barnes M, Ewing-Cobbs L, Fletcher J, Fuchs L. Errors in multi-digit arithmetic and behavioral inattention in children with math difficulties. *Journal of Learning Disabilities*. 2009; 42:356–371. doi:10.1177/0022219409335211. [PubMed: 19380494]
- Rosselli M, Matute E, Pinto N, Ardila A. Memory abilities in children with subtypes of dyscalculia. *Developmental Neuropsychology*. 2006; 30:801–818. doi:10.1207/s15326942dn3003_3. [PubMed: 17083294]
- Ryan JJ, Glass LA, Sullivan DK, Gibson C, Bartels J. PPVT-III alternate forms reliability and stability among inner-city primary school students. *Individual Differences Research*. 2009; 7:70–75.
- Scanlon DM, Gelzheiser LM, Vellutino FR, Schatschneider C, Sweeney JM. Reducing the incidence of early reading difficulties: Professional development for classroom teachers vs. direct interventions for children. *Learning and Individual Differences*. 2008; 18:346–359. doi:10.1016/j.lindif.2008.05.002. [PubMed: 19325710]
- Schatschneider C, Carlson CD, Francis DJ, Foorman BR, Fletcher JM. Relationship of rapid automatized naming and phonological awareness in early reading development: Implications for the double-deficit hypothesis. *Journal of Learning Disabilities*. 2002; 35:245–256. [PubMed: 15493321]
- Shapiro ES, Keller MA, Lutz JG, Santoro LE, Hintze JM. Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment*. 2006; 24:19–35.
- Siegel LS, Ryan EB. The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*. 1989; 60:973–980. [PubMed: 2758890]
- Silver CH, Pennett HD, Black JL, Fair GW, Balise RR. Stability of arithmetic disability subtypes. *Journal of Learning Disabilities*. 1999; 32:108–119. [PubMed: 15499712]
- Speece DL, Case LP. Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology*. 2001; 93:735–740.
- Swanson HL, Beebe-Frankenberger M. The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*. 2004; 96:471–491.
- Swanson HL, Kim K. Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence*. 2007; 35:151–168.
- Swanson J, Schuck S, Mann M, Carlson CD, Hartman K, Sergeant JA, McCleary R. Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN Ratings Scales. Please add Retrieved August. 2005; 12 This is not published. 2006 from <http://www.adhd.net>.
- Temple CM, Sherwood S. Representation and retrieval of arithmetical facts: Developmental difficulties. *Quarterly Journal of Experimental Psychology: Section A*. 2002; 55:733–752.
- Wechsler, D. Wechsler Abbreviated Scale of Intelligence. Psychological Corporation; San Antonio, TX: 1999.
- Wechsler, D. Wechsler Individual Achievement Test Administration Manual. 3rd ed.. Psychological Corporation; San Antonio, TX: 2009.
- Wilkinson, GS. The Wide Range Achievement Test administration manual. Wide Range; Wilmington, DE: 1993.
- Wilkinson, GS.; Robertson, GJ. The Wide Range Achievement Test administration manual. 4th ed.. Psychological Assessment Resources; Lutz, FL: 2006.
- Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock–Johnson III Tests of Achievement. Riverside; Itasca, IL: 2001.
- Wu SS, Meyer ML, Maeda U, Salimpoor V, Tomiyama S, Geary DC, Menon V. Standardized assessment of strategy use and working memory in early mental arithmetic performance. *Developmental Neuropsychology*. 2008; 33:365–393. doi:10.1080/87565640801982445. [PubMed: 18473204]
- Zabel TA, von Thomsen C, Cole C, Martin R, Mahone EM. Reliability concerns in the repeated computerized assessment of attention in children. *Clinical Neuropsychology*. 2009; 23:1213–1231. doi:10.1080/13854040902855358.

Table 1

Diagnostic Grouping at the Initial Time Point.

Criterion	Group	<i>n</i>	%
WRAT-3 Arithmetic	MD	83	57.6
	Control	61	42.4
WJ-III Math Fluency	MD	64	54.2
	Control	55	45.8
Both criteria	MD both	54	45.4
	MD WRAT-3; control WJ-III	9	7.5
	Control WRAT-3; MD WJ-III	10	8.4
	Control both	46	38.7

Note: WRAT-3 = *Wide Range Achievement Test–3rd Edition*; MD = math difficulty; WJ-III = *Woodcock–Johnson Tests of Academic Achievement–III*. A total of 25 students could not be classified under WJ-III criteria given the conditions of the larger study. Percentages within each group sum to 100.

Table 2

Descriptive Statistics on Math Difficulty and Control Groups.

Variable	Category or Scale	MD (<i>n</i> = 83)		No LD (<i>n</i> = 61)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age	Years	9.75	0.82	9.25	0.71
Sex	Female (%)	47.0		52.5	
Ethnicity	African American (%)	52.4		52.5	
	Caucasian (%)	15.9		18.0	
	Hispanic (%)	25.6		26.2	
	Other (%)	6.1		3.3	
Reduced or free lunch	Received total (%)	75.68		51.79	
WASI FSIQ	Standard score	89.77	9.4	101.87	10.7
WRAT-3 Arithmetic	Standard score	86.77	4.6	109.89	7.7
WRAT-3 Reading	Standard score	91.37	12.5	108.23	8.2
WJ-III Math Fluency	Standard score	88.45	12.4	102.39	10.6
SWAN-IV Inattention	-27 to +27	-6.29	9.9	+4.49	13.4
Visual WM	Raw score	7.94	2.4	9.56	3.6
Verbal WM	Raw score	6.33	2.4	7.98	3.1

Note: Values are means and standard deviations, unless otherwise noted. MD = math difficulty; no LD = no learning difficulties; WASI FSIQ = *Wechsler Abbreviated Scale of Intelligence* Full Scale IQ; WRAT-3 = *Wide Range Achievement Test-3rd Edition*; WJ-III = *Woodcock-Johnson Tests of Academic Achievement-III*; SWAN-IV = *Strengths and Weaknesses of ADHD and Normal Behavior-4th Edition*; Visual WM = Visual Working Memory from the *Visuospatial Working Memory Task*; Verbal WM = Verbal Working Memory from the *Categorization Listening Span Task*. Groups are significantly different on all measures, $p < .01$, except for sex and ethnicity.

Table 3

Rate of Categorical Change by Diagnosis and Measure.

Diagnosis	No Change		Change	
	<i>n</i>	%	<i>n</i>	%
WRAT-3 Arithmetic				
No LD	55	90.16	6	9.84
MD	51	61.45	32	38.55
WRAT-3 Arithmetic Severity				
Less severe	28	54.90	23	45.10
More severe	23	71.88	9	28.13
WJ-III Math Fluency				
No LD	48	87.27	7	12.73
MD	40	62.50	24	37.50
WJ-III Math Fluency Severity				
Less severe	15	48.39	16	51.61
More severe	25	75.76	8	24.24

Note: WRAT-3 = *Wide Range Achievement Test–3rd Edition*; no LD = no learning difficulties; MD = math difficulty; WJ-III = *Woodcock–Johnson Tests of Academic Achievement–III*. “More severe” (< 16th percentile) and “less severe” (16th to 32nd percentile) subdivide the groups identified as MD in the section above.

Table 4

Rate of Continuous Change by Diagnosis and Measure.

Diagnosis	Better		Stable		Worse	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
WRAT-3 Arithmetic						
No LD	3	4.92	40	65.57	18	29.51
MD	12	14.46	66	79.52	5	6.02
WRAT-3 Arithmetic Severity						
Less severe	9	17.65	39	76.47	3	5.88
More severe	3	9.38	27	84.38	2	6.25
WJ-III Math Fluency						
No LD	7	12.73	41	74.55	7	12.73
MD	15	23.44	46	71.88	3	4.69
WJ-III Math Fluency Severity						
Less severe	7	22.58	23	74.19	1	3.23
More severe	8	24.24	23	69.70	2	6.06

Note: Continuous change gauged by reliable change index. WRAT-3 = *Wide Range Achievement Test–3rd Edition*; no LD = no learning difficulties; MD = math difficulty; WJ-III = *Woodcock–Johnson Tests of Academic Achievement–III*. “More severe” (< 16th percentile) and “less severe” (16th to 32nd percentile) subdivide the groups identified as MD in the section above.

Table 5

Correlations Among Predictor and Outcome Measures in Full Sample.

Measure	1	2	3	4	5	6	7
1.WRAT-4 Reading							
2.Verbal WM	.227						
3.Visual WM	.306	.481					
4.SWAN-IV Inattention	.323	.299	.368				
5.WJ-III Math Fluency initial	.485	.276	.280	.567			
6.WJ-III Math Fluency final	.406	.243	.329	.541	.706		
7.WRAT-3 Arithmetic initial	.598	.248	.282	.514	.587	.456	
8.WRAT-3 Arithmetic final	.512	.350	.379	.534	.584	.641	.639

Note: Predictor variables administered at initial time point. WRAT-4 = *Wide Range Achievement Test–4th Edition*; Verbal WM = Verbal Working Memory from the *Categorization Listening Span Task*; Visual WM = Visual Working Memory from the *Visuospatial Working Memory Task*; SWAN-IV = *Strengths and Weaknesses of ADHD and Normal Behavior–4th Edition*; WJ-III = *Woodcock–Johnson Tests of Academic Achievement–III*; WRAT-3 = *Wide Range Achievement Test–3rd Edition*. All correlations significant $p < .01$.

Table 6

Regression Statistics for Concurrent and Prospective Math Performance.

	Initial Time Point				Final Time Point				
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	
WRAT-3 Arithmetic initial					0.311	0.086	3.60	.0005	.298
WRAT-3 Reading	0.343	0.070	4.93	<.0001	.363	0.094	1.23	.2227	.095
SWAN-IV Inattention	0.368	0.068	5.38	<.0001	.358	0.274	0.076	3.59	.0005
Verbal WM	0.164	0.118	1.39	.1670	.099	0.289	0.121	2.40	.0178
Visual WM	0.069	0.105	0.66	.5099	.048	0.219	0.107	2.04	.0434
Age ^a	-3.747	1.125	-3.33	.0011	-2.38	-3.652	1.184	-3.08	.0025
WJ-III Math Fluency Initial					0.684	0.108	6.34	<.0001	.578
WRAT-3 Reading	0.319	0.084	3.80	.0002	.313	-0.003	0.101	-0.03	.9767
SWAN-IV Inattention	0.532	0.089	5.97	<.0001	.470	0.200	0.116	1.71	.0875
Verbal WM	0.241	0.139	1.74	.0848	.134	0.044	0.160	0.27	.7854
Visual WM	-0.017	0.127	-0.14	.8915	-0.11	0.242	0.145	1.67	.0973
Age ^a	-3.029	1.388	-2.18	.0312	-.174	-1.452	1.611	-0.90	.3695
Sex ^b	0.484	1.955	0.25	.8047	.0168	-0.965	2.221	-0.43	.6649

Note: *B* = unstandardized beta coefficient; *SE* = standard error; β = standardized beta weight; WRAT-3 = Wide Range Achievement Test-3rd Edition; SWAN-IV = Strengths and Weaknesses of ADHD and Normal Behavior-4th Edition; Verbal WM = Verbal Working Memory from the Categorization Listening Span Task; Visual WM = Visual Working Memory from the Visuospatial Working Memory Task; WJ-III = Woodcock-Johnson Tests of Academic Achievement-III; "initial" refers to math achievement scores from the initial time point appearing as predictors in the final time point models.

^a Age expressed in continuous form (e.g., 10.45 years).

^b Female = 0, male = 1.