Published in final edited form as:

J Learn Disabil. 2016 May ; 49(3): 240–256. doi:10.1177/0022219414538520.

Cognitive Profiles of Mathematical Problem Solving Learning Disability for Different Definitions of Disability

Tammy D. Tolar, PhD¹, Lynn Fuchs, PhD², Jack M. Fletcher, PhD¹, Douglas Fuchs, PhD², and Carol L. Hamlett, MS²

¹University of Houston, TX, USA

²Vanderbilt University, Nashville, TN, USA

Abstract

Three cohorts of third-grade students ($N = 813$) were evaluated on achievement, cognitive abilities, and behavioral attention according to contrasting research traditions in defining math learning disability (LD) status: low achievement versus extremely low achievement and IQ-achievement discrepant versus strictly low-achieving LD. We use methods from these two traditions to form math problem solving LD groups. To evaluate group differences, we used MANOVA-based profile and canonical analyses to control for relations among the outcomes and regression to control for group definition variables. Results suggest that basic arithmetic is the key distinguishing characteristic that separates low-achieving problem solvers (including LD, regardless of definition) from typically achieving students. Word problem solving is the key distinguishing characteristic that separates IQ-achievement-discrepant from strictly low-achieving LD students, favoring the IQ-achievement-discrepant students.

Keywords

mathematics; elementary education; problem solving; learning disability; cognitive

There are at least two approaches to forming groups for research on children with math learning disabilities (LDs): comparing low achievement (LA) to extremely LA (e.g., Murphy, Mazzocco, Hanich, & Early, 2007) and comparing IQ-achievement discrepancy to LA only (Fletcher et al., 1994). Typically groups formed according to these different definitions are compared on a variety of achievement and cognitive measures. The goal of these two approaches is to identify a group of students that is qualitatively distinct and warrants special treatment due to the disability of its members. If such a group exists, then identifying the students and the characteristics that distinguish them may lead to better methods for diagnosis and intervention.

Corresponding Author: Tammy D. Tolar, Department of Educational Psychology, University of Houston, 423 Farish Hall, Houston, TX 77204-5029, USA. ttolar@uh.edu.

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.

In the math LD literature, LA-LD comparisons are more commonly evaluated than IQ-achievement comparisons. Numerical cognition, reading, working memory, processing speed, and visual-spatial processing have been implicated as potential indicators of qualitative differences between groups that differ in degree of low math achievement (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Johnson, Humphrey, Mellard, Woods, & Swanson, 2010; Murphy et al., 2007). Comparisons in this literature have focused more on arithmetical skills than math problem solving, typically comparing LA (< 25th percentile, LA) to extremely LA (< 10th percentile, LD) as indicators of an LD. However, there is some variability in the cut points used to define LD status (Johnson et al., 2010).

The IQ-achievement comparisons have been the subject of a relatively wide body of research, primarily for reading outcomes (Fletcher et al., 1994). For reading LD, meta-analytic evidence based on a range of age groups, reading achievement definition measures, and external validation measures suggests that IQ-achievement-discrepant and LA-only LD students do not differ in achievement or behavior; there are small cognitive differences that can be explained by the relations between the cognitive and group definition variables (Stuebing et al., 2002). These types of comparisons have not received much attention in the math LD literature. Mazzocco and Myers (2003) argue that the IQ-discrepant definition does not reliably discriminate between math disability (MD), borderline, and high-achieving students. However, an argument for evaluating the IQ-discrepant definition is that students who meet this criterion may be qualitatively different from students who are not discrepant but low achieving. If they are qualitatively different, then research on this group may provide some explanatory evidence related to MD. In addition, although states may not require IQ-achievement discrepancy criteria as the sole method for diagnosing an LD (per the Individuals With Disabilities Education Act, 2004, 34 CFR 300.309), many states do allow its use as one of several methods consistent with the Individuals With Disabilities Education Act 2004 (e.g., South Dakota Department of Education, 2013; Texas Education Agency, 2014).

In two of the few studies of MD that compared IQ-achievement-discrepant to LA-only LD students, González and Espinel (1999, 2002) used a test of arithmetic as the definition variable on a group of 7- to 9-year-old students. The IQ-achievement-discrepant students were higher on performance and full IQ but not on verbal IQ. More specifically, the IQ-achievement-discrepant students were statistically higher than the LA group on IQ subtests of comprehension, vocabulary, picture completion, picture arrangement, block design, and object assembly but not arithmetic, similarities, information, and coding. The groups did not, however, differ statistically on working memory as measured by counting span, nor did they differ on word problem solving-performance (e.g., orally administered change, compare, combine, and equalize problems).

In these studies, students were assigned to one of the two comparison groups if their arithmetic achievement scores were below the 25th percentile. They were defined as IQ-achievement discrepant if their arithmetic achievement scores were more than 1 standard deviation lower than their IQ scores; they were defined as IQ-achievement consistent if their achievement scores were within 1 standard deviation of their IQ scores. Yet the achievement cutoff (25th percentile) was not consistent with more recent work that distinguishes students

at or below the 10th percentile from students between the 10th and 25th percentiles (Geary et al., 2007; Murphy et al., 2007). Also, the criteria used for establishing IQ-achievement discrepancy did not account for regression effects that are a consequence of the imperfect relation between IQ and arithmetic achievement (see Cahan, Fono, & Nirel, 2012; Fletcher et al., 1994; Stanovich & Siegel, 1994, for extensive discussions on the consequences of not controlling for regression effects). Finally, it is not surprising that differences emerged on some of the IQ and math measures because they were used to define the groups.

At the same time, both traditions have come under scrutiny in terms of the potential for group differences on achievement or cognitive outcomes to be a consequence of statistical artifacts associated with creating cut points in univariate or multivariate normal distributional spaces—using one or more definitional variables and then comparing the groups on related variables (Branum-Martin, Fletcher, & Stuebing, 2013). For example, if LD definitional groups are formed by dividing students based on levels of arithmetical performance, then the groups formed should differ on skills and abilities that are correlated with arithmetical skill. The magnitudes of expected differences correspond with the strengths of the relations between the skills/abilities and arithmetic performance. Similarly, if these groups are formed based on a combination of actual arithmetical performance and expected performance based on IQ, then the groups should differ on skills and abilities that are correlated with arithmetical skill and IQ. Branum-Martin et al. (2013) demonstrated these psychometric relations in a simulation study. Methods are available to control to some extent for these artifacts (e.g., controlling for the effects of the LD definition variables when evaluating group effects on external variables; see Stanovich & Siegel, 1994), but they have not been consistently applied to math LD in either research tradition.

According to Mazzocco and Myers (2003), the cumulative nature of mathematics as well as qualitative differences in mathematics achievement across grade levels suggests the possibility that MD may manifest at different points and in different ways throughout development. Therefore, it is important to accumulate evidence regarding disability definitions across grades and math skills. One goal of this study was to add to the evidence by evaluating definitions of MD for *math problem solving* among *third-grade students*, which has not received as much research attention as calculation skills and students in lower grades.

We conducted two sets of group comparisons according to the two traditions (low achievement vs. extremely low achievement [LA-LD] and IQ-achievement-discrepant vs. strictly low-achieving LD [IQ-ACH]) among third-grade students evaluated on math-problem-solving achievement. To evaluate potential differences in achievement, we compared the groups on measures of math and reading achievement (math problem solving—nondefinition, word problem solving, arithmetic, computational fluency, word identification, sight word fluency, word reading, and passage comprehension). A math-problem-solving measure that was not used to define the LD groups (nondefinition measure) was included for external validation of the groups. LD groups should differ from non-LD groups but not from each other on a measure that is relatively closely aligned with the problem-solving definition measure. Computational fluency and arithmetic were included because these skills are manifestly required for math problem solving. In addition,

computational fluency has been identified as a persistent problem for students with MDs (Geary, 2004). Word problem solving was included because in some ways it aligns with math problem solving (e.g., problem solving applied to real-world scenarios); however, the numerical manipulation and computational requirements are not as great (at least in the context of this study).

We also compared the groups on cognitive abilities often differentially related to math difficulties (Fuchs et al., 2008) including verbal IQ, nonverbal IQ, language, concept formation, long-term memory, working memory, short-term memory, processing speed, and attentive behavior. The reading achievement measures, cognitive abilities, and attentive behavior are all related to computational fluency and word problem solving among third-grade students (Fuchs et al., 2006; Fuchs et al., 2008). Therefore, it was expected that students who perform poorly on measures of problem solving would also perform poorly on the reading and cognitive/behavioral measures.

In each set of comparisons, we evaluated effect sizes (i.e., univariate group comparisons without controlling for the group definition variables). We expected greater effect sizes for measures that are more highly correlated with the definition variables (e.g., greater group differences on math than reading measures between LD and non-LD groups, greater group differences on measures highly correlated with IQ between IQ-achievement-discrepant and LA-only students).

We also compared the groups using statistical controls to evaluate potentially misleading evidence as a consequence of statistical artifacts common to this type of research (i.e., groups are expected to differ on measures correlated with the group definition measures). We conducted regression analyses including the group and group definition variables as predictors of the external validation variables. This allowed us to evaluate whether there were group differences in achievement and cognitive abilities beyond what would be expected by the group performances on the correlated measures used to define the groups (i.e., problem solving and IQ). Finally, we conducted profile analyses based on MANOVAs to evaluate whether groups differed on achievement or cognitive variables beyond what would be expected given the correlations among the external validation measures (e.g., groups that differ on word reading would be expected to differ to some extent on reading fluency because these measures are correlated). Group differences beyond what would be expected given the correlations among the definition and external validation measures would suggest qualitative differences in the cognitive profiles.

We hypothesized that groups would not differ on most (and perhaps all) achievement and cognitive measures when controlling for the correlations among the measures. However, if there were group differences beyond what would be expected given the relations among the measures, we hypothesized these differences would be in computational fluency and working memory. These have been identified as particularly problematic for children with MDs (Geary, 2004).

Method

Participants

Participants were 813 third-grade students. The mean age of the students at the beginning of third grade was 8.5 years ($SD = 0.4$ years). Of the students, 51% were female, 42% African American, 40% White, 12% Hispanic, 60% on free or reduced-price lunch, and 3% identified with an LD, Attention Deficit Hyperactivity Disorder (ADHD), or a speech, language, or hearing impairment.

These students were selected from three of four cohorts (957 students from consecutive years starting in 2004) on whom data was collected as part of a prospective 4-year study evaluating the effects of mathematics-problem-solving instruction and examining the developmental course and cognitive predictors of mathematics problem solving. One cohort was excluded because some measures evaluated in this study were not administered to that cohort.

Of the 957 students, 137 were excluded because they were missing one or more of the Grade 3 achievement, cognitive, or behavioral measures. The excluded students did not statistically differ from the participants on IQ, 95.2 vs. 97.3, $F(1, 955) = 2.99, p > .05$, or *Woodcock-Johnson III* (WJIII) *Applied Problems*, 100.8 vs. 103.0, $F(1, 953) = 3.28, p > .05$. Seven more students were excluded because performance on one of the measures was extremely low or high relative to their performance on other measures or to the rest of the sample (i.e., 1 student had an IQ score > 150 , 3 students had *Applied Problems* scores more than 30 standard score points lower than their scores the other three times they were measured, 2 students had language standard scores less than 50 although their IQ and problem-solving scores were average, and 1 student had a visual matching score less than 20 standard score points). Based on achievement and cognitive test performance (see Table 1), the 813 students are a representative sample of third-grade students (e.g., *Applied Problems* and IQ scores and standard deviations are consistent with average performance based on test norming samples; see Table 1).

Achievement Measures

Math problem solving—definition—WJIII *Applied Problems* (Woodcock, McGrew, & Mather, 2001a) is used to define LD status in this study. *Applied Problems* is a 60-item test that measures skill in analyzing and solving practical math problems such as counting and telling time or temperature. Items are presented orally, and testing is discontinued after six consecutive errors. Internal consistency and test-retest reliability estimates are $r > .90$ (Alfonso & Flanagan, 2002) and $r = .85$ (McGrew & Woodcock, 2001) for children in the age range of study participants. Age-normed standard scores ($M = 100, SD = 15$) are used for this study.

Math problem solving—nondefinition—Iowa Test of Basic Skills (ITBS): *Problem Solving and Data Interpretation* (Hoover, Dunbar, & Frisbie, 2001) measures skill in solving arithmetical word problems and using data presented in tables and graphs to solve word problems. Internal consistency estimates for Grades 1 to 5 are .83 to .87. The mean grade-

based standard score for third-grade students tested in the spring is 185 (University of Iowa, 2014). The average of standard scores from Forms A and K are used for this study.

Word problem solving—Word-problem-solving scores are based on the sum of scores from *Story Problems* (Jordan & Hanich, 2000) and *Algorithmic Word Problems* (Fuchs, Hamlett, & Powell, 2003). *Story Problems* includes 14 brief story problems involving sums or minuends of 9 or less, with change, combine, compare, and equalize relationships. The tester reads each item aloud; students have 30 s to respond and can ask for rereading(s) as needed. The score is the number of correct answers. Estimated internal consistency is .86 (Fuchs et al., 2008). *Algorithmic Word Problems* includes 9 problems, each of which requires one to four steps. Each problem conforms to a single problem type. Research assistants read aloud each item while students follow along on their own copies of the problems. Problems are scored according to a rubric that assigns points for correct intermediate calculations and final solutions. The maximum score is 19. Estimated internal consistency is .85 (Fuchs et al., 2008).

Arithmetic—The Wide Range Achievement Test-Third Edition Arithmetic Subtest (Wilkinson, 1993) measures skills in counting, number identification, number comparison, and written computations. Internal consistency is high ($r > .90$). Age-normed standard scores ($M = 100$, $SD = 15$) are used for this study.

Computational fluency—Computational fluency scores are based on weighted averages of three assessments: the *Addition Fact Fluency* and *Subtraction Fact Fluency* subtests from the *Grade 3 Math Battery* (Fuchs et al., 2003) and the *Second Grade Test of Computational Fluency* (Fuchs, Hamlett, & Fuchs, 1990). The *Addition* and *Subtraction Fact Subtests* each comprise 25 problems with answers from 0 to 12, and students have 1 min to write answers. Estimated coefficient alpha among third-grade students is .92 (Fuchs et al., 2008). The *Second Grade Test of Computational Fluency* (Fuchs et al., 1990) is a single page of computation items representing the second-grade curriculum: seven single-digit addition items, seven single-digit subtraction items, three double-digit addition items without regrouping, two double-digit addition problems with regrouping, three double-digit subtraction items without regrouping, and three double-digit subtraction items with regrouping. Items are displayed in five rows of 5 items, and students have 3 min to write answers next to or below each item. Estimated coefficient alpha among third-grade students is .94 (Fuchs et al., 2008). Across the three computational fluency measures, the focus is primarily on retrieval of math facts; 11 out of 75 problems across the three measures require double-digit computations, but only 5 of those require regrouping. We computed a problem-correct-per-minute score by averaging a third of the *Computational Fluency* score with the *Addition* and *Subtraction Fact* fluency scores. The maximum score possible is 19.4 problems correct per minute.

Word identification—The Woodcock Reading Mastery Test-Revised, Word Identification (Woodcock, 1998) measures real-world reading ability with 100 words arranged in order of difficulty. Students read words aloud. Testing is discontinued after six consecutive errors at

the end of a page. As per Woodcock, split-half reliability is .98. Age-normed standard scores ($M = 100$, $SD = 15$) are used for this study.

Sight word fluency—Test of Word Reading Efficiency—Sight Word Efficiency (Torgesen, Wagner, & Rashotte, 1999) measures the number of real words students can read in 45 s. As per the test developer, reliability is .95 for 8-year-olds. Age-normed standard scores ($M = 100$, $SD = 15$) are used for this study.

Word reading—The Wide Range Achievement Test-Third Edition, Reading Subtest (Wilkinson, 1993) measures skills in letter identification and word recognition. Internal consistency is high ($> .90$). Age-normed standard scores ($M = 100$, $SD = 15$) are used for this study.

Passage comprehension—Woodcock Reading Mastery Test-Revised, Passage Comprehension (Woodcock, 1998) measures students' ability to identify a picture corresponding to symbols and words and to identify missing words in passages. Split-half reliability exceeds .90. Age-normed standard scores ($M = 100$, $SD = 15$) are used for this study.

Cognitive/Behavioral Measures

IQ—The *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) *Full Scale IQ* score is used to define LD status in this study. The *Full Scale IQ* is based on two subtests: *Vocabulary* and *Matrix Reasoning* (see below). Split-half reliability is .95 for *Full Scale IQ* (Zhu, 1999). Correlations between IQ based on *Vocabulary* and *Matrix Reasoning* alone and IQ based on all four subtests is $r > .93$ (Wechsler, 2003). Age-normed standard scores ($M = 100$, $SD = 15$) are used for this study.

Verbal IQ—WASI *Vocabulary* (Wechsler, 1999) measures expressive vocabulary, verbal knowledge, and foundation of information with 42 items. The first 4 items present pictures; the student identifies the object in the picture. For the remaining items, the tester says a word that the student defines. Responses are awarded a score of 0, 1, or 2, depending on the quality of response. Testing is discontinued after five consecutive scores of 0. As per the test developer, based on a standardization sample representative of the U.S. English-speaking population, split-half reliability coefficients are estimated to be .86 and .88 for 8- and 9-year-olds, and test-retest reliability coefficients are estimated to be .84 for 6- to 11-year-olds. Age-normed T-scores are used in all analyses with expected $M(SD) = 50(10)$ for the representative sample described above.

Nonverbal IQ—WASI *Matrix Reasoning* (Wechsler, 1999) measures nonverbal reasoning with four types of tasks: pattern completion, classification, analogy, and serial reasoning. Examinees look at a matrix from which a section is missing and complete the matrix by selecting among five options. Testing is discontinued after four errors on five consecutive items or after four consecutive errors. As per the test developer, based on a standardization sample representative of the U.S. English-speaking population, split-half reliability coefficients are estimated to be .94 and .93 for 8- and 9-year-olds, and test-retest reliability

coefficients are estimated to be .77 for 6- to 11-year-olds. Age-normed T-scores are used in all analyses with expected $M(SD) = 50(10)$ for the representative sample described above.

Language—Language scores are based on the average of *Test of Language Development Grammatical Closure* (Newcomer & Hammill, 1988) and Woodcock Diagnostic Reading Battery Listening Comprehension (Woodcock, 1997). *Grammatical Closure* measures the ability to recognize, understand, and use English morphological forms. The examiner reads sentences, one at a time; each sentence has a missing word. As reported by the test developer, reliability is .88 for 8-year-olds. *Listening Comprehension* measures the ability to understand sentences or passages. Students supply the word missing from the end of each orally presented sentence or passage. Reliability is .80 at ages 5 to 18. Age-normed standard scores ($M = 100$, $SD = 15$) are used in this study.

Concept formation—WJIII *Concept Formation* (Woodcock, McGrew, & Mather, 2001b) measures the ability of students to identify rules for concepts when shown illustrations of instances and noninstances of the concept. As per the test developer, reliability is .93 for 8-year-olds. Age-normed standard scores ($M = 100$, $SD = 15$) are used in this study.

Long-term memory—WJIII *Retrieval Fluency* (Woodcock & Johnson, 1989) measures long-term memory by asking examinees to recall related items, within categories, for 1 min per category. As reported by the test developer, reliability is .78 for 8-year-olds. Age-normed standard scores ($M = 100$, $SD = 15$) are used in this study.

Working memory—Working Memory Test Battery for Children, Listening Recall (Pickering & Gathercole, 2001) measures students' ability to recall the final words of 1 to 6 sentences after hearing all sentences and determining whether the sentences are true. As per Pickering and Gathercole (2001), test-retest reliability is .93. Age-normed standard scores ($M = 100$, $SD = 15$) are used in this study.

Short-term memory—WJIII *Numbers Reversed* (Woodcock & Johnson, 1989) measures the ability to repeat in reverse order a string of orally presented random numbers. Numbers reversed was evaluated separately from listening recall because evidence suggests that numbers reversed may not represent the same construct as listening recall among older children. Beginning at 8 or 9 years, reverse digit span does not load as highly on a working memory latent factor as listening recall (.50 vs. .69; Gathercole, Pickering, Ambridge, & Wearing, 2004). By 10 to 12 years, when reverse digit span is included in a working memory factor, working memory correlates $> .90$ with a short-term memory factor that includes forward span measures. The two factors based on such a measurement model may represent a single construct. In addition, the different encoding requirements (i.e., numbers vs. words) may cause numbers reversed and listening recall to uniquely discriminate groups defined by levels of math achievement (Dark & Benbow, 1991; Henry, 2001). Numbers reversed is labeled short-term memory in this study because this is how some researchers label it (Rosen & Engle, 1997); however, it may represent more than one construct. As per the test developer, reliability for numbers reversed is .86 for 8-year-olds. Age-normed standard scores ($M = 100$, $SD = 15$) are used in this study.

Processing speed—WJIII *Visual Matching* (Woodcock et al., 2001b) asks examinees to locate and circle two identical numbers that appear in a row of six numbers; examinees have 3 min to complete 60 rows. As per the test developer, reliability is .91 for 8-year-olds. Age-normed standard scores ($M = 100$, $SD = 15$) are used in this study.

Attentive behavior—The *SWAN Rating Scale* (Swanson et al., 2004) is an 18-item teacher rating scale, 9 of which items assess the inattentive behavior criteria for ADHD from the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; American Psychiatric Association, 1994). The items are evaluated on a 7-point scale centered at 0 (range = -3 to 3). For this study the items are reverse scored to represent attentive behavior. The summary score is the sum across the 9 items, ranging from -27 to 27 points. Scores less than -22 would be considered clinically significant for ADHD (derived from 2.48 cutoff for mean item scoring; see Swanson et al., 2004). Internal consistency is estimated to be .98 (Tolar et al., 2012).

Procedure

Test administration—All math achievement tests except WJIII *Applied Problems* (the LD definition measure) and ITBS *Problem Solving* were administered in September in whole-class sessions. ITBS *Problem Solving* was administered in whole-class sessions in early March. WJIII *Applied Problems*, all reading achievement tests except WJIII *Passage Comprehension*, and all cognitive tests were administered during September and October in individual test sessions by trained research assistants according to standard instructions. WJIII *Passage Comprehension* was administered in individual sessions in late March and early April. The individual test sessions were audio-taped, and 17.9% of tapes, distributed equally across testers, were selected randomly for accuracy checks by an independent scorer. Agreement estimates exceeded 98.8%. In October, teachers completed the *SWAN Rating Scale* on each student.

LD group classifications—Students were classified for two types of comparisons: (a) LA versus LD (LA-LD) comparisons and (b) IQ-ACH comparisons.

LA-LD comparison groups—We used sample-based percentiles to classify students into learning disabled, low achieving, and no LD. The sample of Grade 3 students in this study was substantially larger than the one used for norming WJIII *Applied Problems* (813 vs. 490; McGrew, Schrank, & Woodcock, 2007) and differed in demographic makeup (e.g., 42% vs. 15% African American). The sample-based cutoffs for the 10th and 25th percentile were 87 and 94, respectively, whereas the test norm cutoffs are 80 and 90, suggesting the study sample has greater problem-solving skills than the test-norming sample. For this study, if students scored less than or equal to a standard score of 87 on math problem solving-definition they were classified as learning disabled. If they scored greater than 87 but less than or equal to 94 they were classified as low achievers. If they scored greater than 94 they were classified as having no LD. Based on these definitions, 74 students were classified as learning disabled, 134 as low achieving, and 605 as not learning disabled.

IQ-ACH comparison groups—We used the methods described in Fletcher et al. (1994) to identify students with discrepancies between their math-problem-solving performance and their predicted performance based on IQ. The sample-based correlation between IQ and WJIII *Applied Problems* ($r = .60$) was used to calculate predicted math-problem-solving scores and standard errors of prediction. The correlation between IQ and math reasoning (WJIII *Applied Problems*) among the test-norming samples ranged from .61 to .73 depending on the IQ measure and age group (6-8 and 9-13 years, correlations are lower among the younger children; McGrew et al., 2007). This suggests that math-problem-solving achievement may be slightly less related to IQ among the study students. For this study, students whose actual math-problem-solving scores were less than or equal to 1.5 standard errors of prediction below predicted scores were classified as learning disabled based on an IQ-achievement discrepancy. Students who were not classified as IQ-achievement discrepant but scored less than or equal to 87 (10th percentile) on math problem solving were classified as learning disabled based on achievement alone. According to these definitions, 42 students were classified as learning disabled due to an IQ-achievement discrepancy, 42 as learning disabled based on achievement alone, and 729 as not learning disabled. Among the IQ-achievement-discrepant students, 32 (76%) also met criteria for math-problem-solving LD based on achievement alone.

Analytic procedures—For each set of comparison groups, three types of analyses were conducted: profile, canonical, and regression. All analyses were conducted separately on two sets of outcomes: achievement and cognitive/behavioral. The same achievement measures were used in both sets of group comparisons; however, for the cognitive/behavioral measures, verbal and nonverbal IQ were included in the LA-LD but not the IQ-ACH comparisons because they were nondefinition variables for the former but not the latter.

Profile analyses—Profile analysis is based on MANOVA and typically involves tests for shape, level, and flatness. Shape effects (group by outcome interactions) indicate different profiles (i.e., qualitative differences) and are the focus of this study. In the event of no shape effect, level or group main effects indicate group differences across outcomes. Flatness is indicated by main effects of outcome (i.e., across groups, scores are higher on one outcome than another). Any potential flatness effects were removed by using z-scores (calculated on the full sample of 813 students) in the MANOVA analyses. For the profile analyses, we included the no-LD as well as the LD groups from each set of comparison groups.

Canonical analyses—Canonical analysis is derived from MANOVA, which computes a linear function of the outcome variables that maximizes the difference between the groups being evaluated. It permits interpretation of the discriminant functions based on the weightings of the variables making up the discriminant function (Huberty & Olejnik, 2006). We evaluated both the standardized discriminant function coefficients, which represent the relative weights of the variables in the discriminant functions, and the canonical correlations, which represent the absolute relations between the discriminant functions and the outcome variables. For this set of analyses we included only the LD groups (i.e., LA-LD: low achieving vs. learning disabled; IQ-ACH: learning disabled based on IQ discrepancy vs. learning disabled based on LA alone).

Regression analyses—Regression analyses were conducted for each outcome. Each regression model included the LD definition variables: math problem solving for the LA-LD comparison groups and math problem solving and IQ for the IQ-ACH comparison groups (see Stanovich & Siegel, 1994, for detailed explanation and application of this method to reading disabilities). Groups formed from the division of continuous, normally distributed measures are more likely to differ on measures that are more highly correlated with the definition measures. This is a statistical artifact. By controlling for the definition variables, we are evaluating whether individuals at similar levels of problem solving and IQ performance differ by LD status on the outcome variables.

All regression models were first evaluated with just the definition variables to determine if there were nonlinearities. Nonlinearities were identified between math problem solving-nondefinition and math problem solving-definition, word problem solving and IQ, and word identification and IQ. Therefore, for the LA-LD comparison group regression analysis involving math problem solving-nondefinition, a squared term for math problem solving-definition was included in the model. For IQ-ACH comparison group analyses involving word problem solving and word identification, a squared term for IQ was included in the models.

Two contrast variables were evaluated in the regression models: The no-LD groups versus the LA-LD or IQ-ACH groups and LA versus LD or IQ versus ACH groups. This allowed us to evaluate whether there is an additional effect of LD status on the outcome variable beyond general low mathematical-problem-solving status.

Results

Evaluation of Cohorts

The three cohorts of students did not statistically differ in group assignments, LA-LD, $\chi^2(4) = 1.71, p > .05$; IQ-ACH, $\chi^2(4) = 1.59, p > .05$. MANOVA was used to compare the cohorts on all measures used in the study. There was a significant measure by group interaction, $F(32, 12960) = 2.49, p < .05$. However, the significant effect is likely due to the large group sizes ($n > 258$ per group) and the large number of variables. The differences did not appear to be practically significant. One effect size was .39 (sight word fluency, standard scores of 105.1 and 100.5 for Cohorts 3 and 4, respectively). The rest of the effect sizes were less than .30. There was no systematic pattern in the cohort differences. For these reasons, all subsequent analyses were collapsed across cohorts.

Effect Sizes

Raw effect sizes were calculated to evaluate the magnitude of group differences without controlling for the relations between the group definition variables and the achievement, cognitive, and behavioral variables.

Definition variables—The largest differences between groups were, as expected, in the definition variables (see Table 2). For the LA-LD comparison groups, the effect sizes (i.e., number of standard deviations difference) for math problem solving-definition were 1.36 (no-LD vs. LA), 2.12 (no-LD vs. LD), and 0.76 (LA vs. LD). For the IQ-ACH comparison

groups, there was a small difference in IQ between no-LD students and IQ-achievement-discrepant LD students (effect size = 0.25; see Table 2) and a small difference in math-problem-solving achievement (definition) between IQ-achievement-discrepant and LA-only (ACH) LD students (effect size = -0.18). Otherwise, for the IQ-ACH comparison groups, the effect sizes were large when comparing no-LD students to IQ and ACH students on math problem solving-definition (1.87 and 1.69; see Table 2), no-LD students to ACH students on IQ (1.22), and IQ to ACH students on IQ (0.97).

Nondefinition variables—Overall, there were large differences (effect sizes > 0.60; see Table 2) in achievement, cognitive abilities, and attentive behavior when comparing no-LD students to low-achieving or LD students (regardless of the way the LD groups were defined) with two general exceptions. First, the differences in memory (long term, working, and short term) between no-LD and low-achieving students (LA-LD comparison groups) were more moderate than differences on other achievement and cognitive measures (effect sizes = 0.54, 0.55, and 0.60). Second, the differences in cognitive abilities were small to moderate (effect sizes 0.19-0.49) between no-LD and IQ-discrepant students except for concept formation (0.65), processing speed (0.80), and attentive behavior (0.63).

In general, the differences in achievement, cognitive abilities, and attentive behavior were low to moderate (see Table 2) between low-achieving and LD students (LA-LD definitional groups) and between IQ-discrepant LD and LA-only LD students (IQ-ACH definitional groups). The largest effect size for the LA versus LD comparison was in arithmetic (0.69) and for IQ versus ACH comparison was in working memory (0.58).

Profile Analyses

LA-LD comparison groups—There were LA-LD group differences in profile shape for the achievement but not the cognitive/behavioral outcomes as indicated by a statistically significant achievement by group interaction, $F(14, 5670) = 4.16, p < .05$, but nonsignificant cognitive/behavioral outcome by group interaction, $F(16, 6480) = 1.18, p > .05$. The achievement shape differences appear to be primarily between the LA and LD groups. Based on post hoc analyses, LA students scored significantly higher than LD students on all math measures, word reading, and passage comprehension ($p < .05$) but not word identification or sight word reading fluency (see Table 3). The no-LD students scored significantly higher than both the LA and LD students on all achievement measures.

There were level differences in cognitive abilities between LA-LD groups as indicated by a statistically significant main effect of group, $F(2, 810) = 149.46, p < .05$. These differences appear to be primarily between the no-LD students and the two lower achieving groups (LA and LD). Based on post hoc comparisons, the no-LD students scored statistically higher than both low-achieving groups on all cognitive/behavioral measures ($p < .05$; see Table 3) whereas the LA and LD students did not statistically differ on any measure with the exception of concept formation and working memory.

The LA-LD shape effects in the achievement measures and level effects in the cognitive/behavioral measures can be seen in Figure 1. There is a slightly larger gap (approximately .5 or greater *z* score) among the math than the reading measures between the LA and LD

students whereas there is a relatively small gap (approximately .25 or less *z* score) across all cognitive/behavioral measures between the two groups of students. The difference between the no-LD and lower achieving groups was consistently large (approximately .75 or greater *z* score) across all achievement and cognitive/behavioral measures.

IQ-ACH comparison groups—Similar to the LA-LD group comparisons, there were differences between the IQ-ACH comparison groups in profile shapes for the achievement but not the cognitive/behavioral outcomes as indicated by a statistically significant achievement by group interaction, $F(14, 5670) = 3.58, p < .05$, but nonsignificant cognitive/behavioral outcome by group interaction, $F(12, 4860) = 1.33, p > .05$. Based on post hoc analyses, IQ-discrepant LD students scored significantly higher than low-achieving-only LD students on word problem solving ($p < .05$) but none of the other achievement measures (see Table 3). The no-LD students scored significantly higher than both the IQ-discrepant and low-achieving-only LD students on all achievement measures.

There were level differences in cognitive abilities across the IQ-ACH groups as indicated by a statistically significant main effect of group, $F(2, 810) = 46.33, p < .05$. Similar to the LA-LD groups, these differences appear to be primarily between the no-LD students and the two LD groups. Based on post hoc comparisons, the no-LD students scored statistically higher than both LD groups on all cognitive/behavioral measures ($p < .05$; see Table 3) whereas the IQ-discrepant and low-achieving-only LD students did not differ statistically on any measure with the exception of working memory.

The IQ-ACH shape effects in the achievement measures and level effects in the cognitive/behavioral measures can be seen in Figure 2. For the achievement measures, there was a moderate gap ($> .5$ *z* score) between the IQ-discrepant and low-achieving-only LD students in word problem solving but virtually no gap in most of the other measures with the exception of a small gap ($< .5$ *z* score) in math problem solving-nondefinition. The gap between no-LD and LD students was consistently large (approximately .75 or greater *z* score) across achievement measures. For the cognitive/behavioral measures, there was a moderate to large gap between the no-LD and LD groups across measures. The gap was generally small to nonexistent across measures between the IQ-discrepant and low-achieving-only LD students with the exception of working memory. However, this noticeable difference in working memory favoring IQ-discrepant LD students did not rise to the level of a statistically significant shape difference in the multivariate context.

Canonical Analyses

LA-LD comparison groups (excluding no LD)—For the achievement measures, the canonical analysis results were consistent with the shape effects in the profile analyses. The magnitudes of the standardized discriminant function and correlation coefficients were generally greater among the math ($r_s = .53$ to $.78$; see Table 3) than reading ($r_s = .22$ to $.48$) outcomes and greater among word reading ($r = .39$) and passage comprehension ($r = .48$) than word identification and sight word efficiency ($r = .22$ for both). The canonical correlation was highest for general arithmetic (.78 compared to the next highest, .59, for math problem solving-nondefinition; see Table 3), suggesting that general arithmetic is the

most discriminating of the achievement measures between LA and LD students. For the cognitive/behavioral variables, the canonical correlations were generally low to moderate with the exception of concept formation ($r = .79$; see Table 3), which appears to be the most discriminating of the cognitive measures between LA and LD students.

IQ-ACH comparison groups (excluding no LD)—For the achievement measures, the canonical analysis results were mostly consistent with the shape effects in the profile analyses. However, whereas the IQ-discrepant and low-achieving-only LD students differed only on word problem solving in the profile analysis, based on the canonical correlations, math problem solving-nondefinition, word problem solving, and computational fluency appear to be similarly discriminating between the two groups ($r_s = .47, .48$, and $-.42$, respectively; see Table 4). The magnitude of the canonical correlations for the reading measures and general arithmetic were substantially lower ($r_s = -.15$ to $.24$). For the cognitive/behavioral variables, the canonical correlations were generally low to moderate with the exception of language and working memory ($r = .69$ and $.68$; see Table 4), which appears to be similarly discriminating between LA and LD students.

Regression Analyses

LA-LD comparison groups—In addition to the canonical analysis, the regression analyses provide converging evidence that general arithmetic is the key discriminator between LA and LD students. It also discriminates between no LD and generally LA in math problem solving. When controlling for math problem solving-definition, arithmetic was the only achievement variable on which no-LD students differed statistically from all low-achieving students (no LD vs. LA/LD: $t(809) = -2.20, p < .05$) and LA students were statistically different from LD students (LA vs. LD: $t(809) = -2.37, p < .05$; see Table 3 for all other comparisons). The only other statistical differences when controlling for math problem solving-definition were between no LD and the low-achieving students (LA/LD) on sight word fluency, $t(809) = -2.41, p < .05$, and general reading, $t(809) = -2.01, p < .05$. When controlling for the group definition variable, math problem solving, there were no statistical differences between the no-LD and low-achieving students or between LA and LD students on any of the cognitive/behavioral measures (see Table 3).

IQ-ACH comparison groups—The regression analyses results were consistent with both the profile and canonical analyses supporting word problem solving as the main discriminator between IQ-discrepant and low-achieving-only LD students. It was the only achievement measure for which the two groups differed statistically when controlling for both math problem solving-definition and IQ, $t(809) = -2.31, p < .05$. The only other statistical difference in achievement when controlling for math problem solving-definition and IQ was between no LD and LD (both IQ discrepant and low achieving only) on general arithmetic, $t(809) = -2.04, p < .05$ (see Table 4 for all other comparisons). For the cognitive/behavioral measures, the only statistical difference between groups when controlling for math problem solving-definition and IQ was between IQ-discrepant and low-achieving-only LD on working memory, $t(809) = -2.01, p < .05$ (see Table 4 for all other comparisons). This was consistent with the canonical analyses but inconsistent with the profile analyses,

which suggest that when accounting for all cognitive/behavior variables at once, there is no statistical difference (i.e., shape effect) between the groups.

Discussion

Our goal was to evaluate whether groups formed according to different definitions of math-problem-solving LD differ in achievement and cognitive/behavioral outcomes. We evaluated group differences according to two research traditions: (a) low achieving (< 25th percentile) versus LD (< 10th percentile) and (b) IQ-achievement-discrepant versus low-achieving-only LD. We based our evaluations on three statistical methods: profile and canonical analyses to control for the relations among the nondefinition variables and regression analyses to control for the relations between the definition and nondefinition variables.

The key finding was that regardless of how the groups are defined, there are shape differences in the profiles among the achievement but not the cognitive/behavioral measures. When the groups are defined strictly by math-problem-solving achievement level (LA vs. LD), they differ only in basic arithmetic skills. When the groups are defined by IQ discrepancy (IQ-discrepant vs. low-achieving LD only), they differ only in word problem solving. Working memory is the only cognitive/behavioral variable on which the IQ-discrepant and low-achieving-only students differ (even when controlling for the group definition variables). However, this difference does not appear to be distinct in the context of other cognitive/behavioral variables due to the lack of a shape effect between the groups.

Although there are moderate differences in other achievement and cognitive abilities between LD groups (regardless of how the groups are defined), most of these differences can be explained by the relations among the nondefinition measures or between the definition and non-definition measures. That is, math problem solving is correlated with most measures of achievement, cognitive abilities, and attentive behavior, and most of these measures are correlated with each other. Therefore, students who tend to do well in math problem solving tend to do well on all of these measures, and students who are poor math problem solvers tend to do poorly on all of these measures. This pattern is also true for IQ.

Even so, there are two measures in which there are greater differences in groups than would be expected given the relations among math problem solving, IQ, and the other measures of achievement, cognitive abilities, and attentive behavior. These measures are basic arithmetic and word problem solving.

For the LA-LD comparisons, the univariate differences across math measures and in concept formation and working memory are consistent with other studies based on this method of defining groups (Geary et al., 2007; Murphy et al., 2007). Our study demonstrates that for most of these measures with the exception of general arithmetic, the differences between the groups are explained by the difference in their problem-solving ability. However, it is not clear why basic arithmetic is unique among the other math skills in distinguishing these two groups. One possible explanation is that math problem solving-nondefinition and word problem solving are relatively similar to WJIII *Applied Problems* in that they emphasize language comprehension and math reasoning over computation. It is therefore not

unexpected that when controlling for performance on *Applied Problems*, there is no difference between groups in the other problem-solving measures. The relatively low-level computational skills required for computational fluency may also be why level differences on *Applied Problems* (i.e., group definition) explain any differences in computational fluency. Arithmetic (i.e., the *Wide Range Achievement Test, Arithmetic* subtest) requires decontextualized counting, identifying numbers, number comparison, and written computations, which is a broader range of skills than those required for computational fluency but consistent with those of *Applied Problems*. It may be that the ability to work with numbers in a variety of ways in the absence of context is what uniquely (relative to all other achievement and cognitive measures) distinguishes different levels of problem-solving ability.

There are several numerical abilities related to third-grade math achievement that were not evaluated in this study, including counting knowledge (Geary, 2004), estimation (Jordan & Hanich, 2003), number line estimation (Geary, Bailey, & Hoard, 2009; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013), number sets (Geary et al., 2009), digit comparison (Sasanguie et al., 2013), and the approximate number system (Halberda, Mazzocco, & Feigenson, 2008). These numerical abilities are sometimes grouped under the term “number sense.” It has been suggested that number sense is to math as phonological awareness is to reading and as such is an important indicator of future math performance (Gersten & Chard, 1999). Nonsymbolic number line estimation and number sets may be especially relevant to third-grade math achievement (Sasanguie et al., 2013). It is not clear if these number sense abilities are captured by the group differences in arithmetic or if they would uniquely explain (compared to arithmetic) problem-solving differences in the groups. It is possible that the inclusion of measures of nonsymbolic number line estimation and number sets in the multivariate analysis would explain the group differences in general arithmetic.

For the IQ-discrepant/low-achieving-only comparisons, there were univariate differences only in word problem solving and working memory, even when controlling for the group definition variables. In terms of the univariate results, the IQ-ACH group differences in working memory and word problem solving were not consistent with other evidence (González & Espinel, 1999, 2002). There are several differences between this study and the González and Espinel studies that may account for the different results. González and Espinel used an arithmetic measure to define LD groups, whereas we used math problem solving as the definition variable. Their working memory task was a dot span task requiring counting dots and remembering numbers, whereas ours was a listening recall task requiring sentence comprehension and remembering words. The former is a numerical working memory task, whereas the latter is a working memory task involving language. It should be noted that in González and Espinel’s study, the IQ-achievement-discrepant students were higher on vocabulary (i.e., comparable to differences in listening recall in our study) than the strictly low-achieving students. Also, in our study, there was no difference between the groups on the numbers reversed short-term memory task (i.e., comparable to the lack of difference in numerical working memory in the González and Espinel studies). Finally, the word-problem-solving tasks were very similar across studies, both involving orally presented story problems involving relatively simple computations.

The working memory measure used in this study (i.e., *Listening Recall*) requires language comprehension as well as short-term memory for words. The IQ-discrepant and low-achieving-only LD group difference in working memory is not unique among the cognitive/behavioral measures given the lack of a shape effect. Yet it is possible the language and memory requirements shared across the cognitive measures but fully encapsulated in the working memory measure explains the group difference in word problem solving (which is unique among the math measures given the shape effect in achievement outcomes). That is, IQ-achievement-discrepant students who are very low achieving (10th percentile) have a greater than expected advantage in language and short-term memory for words (hence their greater IQ) over very-low-achieving students who are not IQ discrepant. This gives IQ-discrepant students an advantage over nondiscrepant students in a word-problem-solving task that requires greater oral language than computational resources (e.g., the word-problem-solving tasks with simple computations) but not on problem-solving tasks that are more demanding of computational resources (e.g., the math-problem-solving definition variable that is highly correlated with arithmetic). However, the advantage that IQ-discrepant students have over nondiscrepant students in word problem solving may not be functionally meaningful given both groups' extremely low performance relative to non-LD students.

In comparisons of IQ-achievement-discrepant and low-achieving-only students, the results from this study are consistent with other literature focused on comparisons between children with MDs but no reading disabilities (MD) and those with both math and reading disabilities (MD/RD) (Jordan & Hanich, 2003; Jordan, Hanich, & Kaplan, 2003b; Jordan & Montani, 1997). The MD children from these studies may be somewhat comparable to the IQ-achievement-discrepant students (who do not differ from no-LD students on IQ), and the MD/RD children are somewhat comparable to the low-achieving-only students (who have lower IQs than both IQ-achievement-discrepant and no-LD students) based on average levels of IQ. Although MD and MD/RD children are equally poor at retrieving math facts under timed conditions, the MD children perform better than MD/RD children when there is no time constraint (Jordan & Montani, 1997). In untimed conditions, MD children are able to use backup counting strategies to outperform MD/RD children. Math fact mastery (i.e., efficient retrieval from long-term memory) appears to be independent of language and reading abilities (Jordan, Hanich, & Kaplan, 2003a). MD children who have poor math fact mastery but average language and reading abilities may be able to leverage language abilities in applying backup strategies to perform simple calculations, although these language abilities do not compensate for nonverbal number sense abilities (e.g., estimation; Jordan & Hanich, 2003). In the case of number sense abilities, MD children do not differ from MD/RD. In the context of this study, MD children would be expected to outperform MD/RD children on untimed math tasks where language abilities could be used to compensate for poor math fact mastery and poor number sense abilities (similar to IQ vs. ACH students on word problem solving) but perform comparably to MD/RD children on arithmetical tasks that require timed retrieval of math facts or number sense abilities (similar to IQ vs. ACH students on computational fluency, general arithmetic, and applied problem solving).

Practical Implications

It is possible that the IQ-discrepant/-nondiscrepant distinction may warrant different types of intervention, perhaps focusing on working memory and related cognitive resources, for example, helping IQ-discrepant students identify and use problem-solving strategies that leverage their relative advantage (compared to low-achieving-only LD students) for processing and holding verbal information in short-term memory. Future studies establishing the effectiveness of intervening in this way are needed to evaluate the potential of investing resources into this type of individualized instruction. In making these distinctions, it should be recognized that there is overlap between the groups in working memory and it is best to directly assess working memory regardless of IQ discrepancy. A stronger argument could be made for focusing on the practical implications of arithmetic as a distinguishing characteristic between math-problem-solving-LD and non-LD students.

Students who are low achieving in math problem solving, whether they meet criteria for LD or not, appear to have a particular problem with general arithmetic, although it is not clear if the problem is with procedural computational abilities, number sense, or a combination of the two. Targeted intervention that focuses on improving these students' general arithmetic including procedural skills and knowledge about relations between quantities (i.e., number sense) may be a necessary first step to improving math-problem-solving skills.

Limitations

A criticism of research evaluating cognitive factors related to LD is that we do not know if they are causal, or a consequence of the LD, or a covariate (Büttner & Hasselhorn, 2011). This study does not address these possibilities. The results from this study should be treated as preliminary evidence for future studies that use other methods (e.g., experimental, longitudinal) that evaluate whether the potential markers of problem-solving LD identified by this study *predict* problem-solving performance or can be manipulated to improve performance.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant RO1 HD46154 and Core Grant HD15052 to Vanderbilt University, and Grant K99HD061689 to the University of Houston). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health.

References

- Alfonso, VC.; Flanagan, DP. Comparative features of comprehensive achievement batteries. Riverside; Ithaca, IL: 2002. Woodcock-Johnson III Assessment Service bulletin No. 5
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed.. Author; Washington, DC: 1994.
- Branum-Martin L, Fletcher JM, Stuebing KK. Classification and identification of reading and math disabilities: The special case of comorbidity. *Journal of Learning Disabilities*. 2013; 46(6):490–499. doi: 10.1177/0022219412468767. [PubMed: 23232442]

- Büttner G, Hasselhorn M. Learning disabilities: Debates on definitions, causes, subtypes, and responses. *International Journal of Disability, Development & Education*. 2011; 58(1):75–87. doi: 10.1080/1034912x.2011.548476.
- Cahan S, Fono D, Nirel R. The regression-based discrepancy definition of learning disability: A critical appraisal. *Journal of Learning Disabilities*. 2012; 45(2):170–178. [PubMed: 20375287]
- Dark VJ, Benbow CP. Differential enhancement of working memory with mathematical versus verbal precocity. *Journal of Educational Psychology*. 1991; 83(1):48–60.
- Fletcher JM, Shaywitz SE, Shankweiler DP, Katz L, Liberman IY, Stuebing KK, Shaywitz BA. Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology*. 1994; 86:6–23.
- Fuchs LS, Fuchs D, Compton DL, Powell SR, Seethaler PM, Capizzi AM, 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 K, 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]
- Fuchs, LS.; Hamlett, CL.; Fuchs, D. Curriculum-based math computation and concepts/applications. Vanderbilt University; Nashville, TN: 1990. Unpublished manuscript
- Fuchs, LS.; Hamlett, CL.; Powell, SR. Grade 3 math battery. 2003. Available from L. S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203
- 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]
- Geary DC. Mathematics and learning disabilities. *Journal of Learning Disabilities*. 2004; 37:4–15. [PubMed: 15493463]
- Geary DC, Bailey DH, Hoard MK. Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The number sets test. *Journal of Psychoeducational Assessment OnlineFirst*. 2009; 27:265–279. doi: 10.1177/0734282908330592.
- 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]
- Gersten R, Chard D. Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education*. 1999; 33(1):18.
- González JEJ, Espinel AIG. Is IQ-achievement discrepancy relevant in the definition of arithmetic learning disabilities? *Learning Disability Quarterly*. 1999; 22:291–301.
- González JEJ, Espinel AIG. Strategy choice in solving arithmetic word problems: Are there differences between students with learning disabilities, G-V poor performance and typical achievement students? *Learning Disability Quarterly*. 2002; 25:113–122. doi: 10.2307/1511278.
- Halberda J, Mazocco MLMM, Feigenson L. Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*. 2008; 455(7213):665–668. [PubMed: 18776888]
- Henry LA. How does the severity of a learning disability affect working memory performance? *Memory*. 2001; 9(4/6):233–247. doi: 10.1080/09658210042000085.
- Hoover, HD.; Dunbar, SB.; Frisbie, DA. Iowa test of basic skills. Riverside; Rolling Meadows, IL: 2001.
- Huberty, CJ.; Olejnik, S. Applied discriminant analysis. 2nd ed.. Wiley; New York, NY: 2006.
- Johnson ES, Humphrey M, Mellard DF, Woods K, Swanson HL. Cognitive processing deficits and students with specific learning disabilities: A selective meta-analysis of the literature. *Learning Disability Quarterly*. 2010; 33(1):3–18.
- Jordan NC, Hanich LB. Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities*. 2000; 33(6):567–578. [PubMed: 15495398]
- Jordan NC, Hanich LB. Characteristics of children with moderate mathematics deficiencies: A longitudinal perspective. *Learning Disabilities Research & Practice*. 2003; 18(4):213. doi: 10.1111/1540-5826.00076.

- Jordan NC, Hanich LB, Kaplan D. Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*. 2003a; 85:103–119. [PubMed: 12799164]
- 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*. 2003b; 74:834–850. [PubMed: 12795393]
- Jordan NC, Montani TO. Cognitive arithmetic and problem solving: A comparison of children with specific and general mathematics difficulties. *Journal of Learning Disabilities*. 1997; 30:624–634. [PubMed: 9364900]
- Mazzocco MM, Myers GF. Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Ann Dyslexia*. 2003; 53(1):218–253. doi: 10.1007/s11881-003-0011-7. [PubMed: 19750132]
- McGrew, KS.; Schrank, FA.; Woodcock, RW. *Woodcock-Johnson III normative update* (technical manual). Riverside; Rolling Meadows, IL: 2007.
- McGrew, KS.; Woodcock, RW. *Woodcock-Johnson III technical manual*. Riverside; Itasca, IL: 2001.
- 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]
- Newcomer, PL.; Hammill, DD. *Test of language development*. Pro-Ed; Austin, TX: 1988.
- Pickering, S.; Gathercole, S. *Working memory test battery for children*. Psychological; London, UK: 2001.
- Rosen VM, Engle RW. Forward and backward serial recall. *Intelligence*. 1997; 25(1):37–47. doi: [http://dx.doi.org/10.1016/S0160-2896\(97\)90006-4](http://dx.doi.org/10.1016/S0160-2896(97)90006-4).
- Sasanguie D, Göbel SM, Moll K, Smets K, Reynvoet B. Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*. 2013; 114(3):418–431. doi: <http://dx.doi.org/10.1016/j.jecp.2012.10.012>. [PubMed: 23270796]
- South Dakota Department of Education. Eligibility guide for special education. 2013. Retrieved from <http://doe.sd.gov/oess/documents/Eligibilt.pdf>
- Stanovich KE, Siegel LS. Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable difference model. *Journal of Educational Psychology*. 1994; 86:24–53.
- Stuebing KK, Fletcher JM, LeDoux JM, Lyon GR, Shaywitz SE, Shaywitz BA. Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal*. 2002; 39:469–518.
- Swanson, J.; Schuck, S.; Mann, M.; Carlson, C.; Hartman, K.; Sergeant, J.; McCleary, R. Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and SWAN rating scales. 2004. Retrieved from <http://www.adhd.net>
- Texas Education Agency. Evaluation of learning disability (LD) eligibility. 2014. Retrieved from <http://www.tea.state.tx.us/index2.aspx?id=2147500368>
- Tolar TD, Fuchs L, Cirino PT, Fuchs D, Hamlett CL, Fletcher JM. Predicting development of mathematical word problems solving across intermediate grades. *Journal of Educational Psychology*. 2012; 104:1083–1093. doi: 10.1037/a0029020. [PubMed: 23325985]
- Torgesen, JK.; Wagner, R.; Rashotte, C. *Test of word reading efficiency*. Pro-Ed; Austin, TX: 1999.
- University of Iowa. Interpreting results from the Iowa assessments™. 2014. Retrieved from http://itp.education.uiowa.edu/ia/documents/IA_Interpreting_Results.pdf
- Wechsler, D. *Wechsler abbreviated scale of intelligence*. Psychological; San Antonio, TX: 1999.
- Wechsler, D. *Manual for the Wechsler abbreviated scale of intelligence*. fourth edition. Psychological; San Antonio, TX: 2003.
- Wilkinson, GS. *The wide range achievement test administration manual*. Wide Range; Wilmington, DE: 1993.
- Woodcock, RW. *Woodcock diagnostic reading battery*. Riverside; Itasca, IL: 1997.

- Woodcock, RW. Woodcock reading mastery test-revised. American Guidance Service; Circle Pines, MN: 1998.
- Woodcock, RW.; Johnson, MB. Woodcock-Johnson psycho-educational battery-revised. DLM Teaching Resources; Allen, TX: 1989.
- Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III tests of achievement. Riverside; Itasca, IL: 2001a.
- Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III tests of cognitive abilities. Riverside; Itasca, IL: 2001b.
- Zhu, J. Wechsler abbreviated scale of intelligence manual. Author; San Antonio, TX: 1999.

Figure 1.

Achievement (a) and cognitive (b) profiles of LA-LD definitional groups.

Note. LA-LD = low achievement versus extremely low achievement; mps-nd = math problem solving-nondefinition; wps = word problem solving; arith = arithmetic; comp = computational fluency; wi = word identification; sw = sight word fluency; rd = word reading; pc = passage comprehension; viq = verbal IQ; nviq = nonverbal IQ; lang = language; cf = concept formation; ltm = long-term memory; wm = working memory; stm = short-term memory; ps = processing speed; att = attentive behavior.

Figure 2.

Achievement (a) and cognitive (b) profiles of IQ-ACH definitional groups.

Note. IQ-ACH group = IQ-achievement-discrepant versus strictly low-achieving learning disability; mps-nd = math problem solving-nondefinition; wps = word problem solving; arith = arithmetic; comp = computational fluency; wi = word identification; sw = sight word fluency; rd = word reading; pc = passage comprehension; viq = verbal IQ; nviq = nonverbal IQ; lang = language; cf = concept formation; ltm = long-term memory; wm = working memory; stm = short-term memory; ps = processing speed; att = attentive behavior.

Table 1Means, Standard Deviations, and Correlations With Learning Disability Definition Measures ($N = 813$).

Measure	<i>M</i>	<i>SD</i>	<i>r</i>	
			MPS-def	IQ
Learning disability definition				
Math problem solving–definition	103.2	12.5		
IQ	97.3	13.2	.60	
Achievement				
Math problem solving–nondefinition	184.3	19.1	.69	.59
Word problem solving	15.0	6.6	.66	.55
Arithmetic	97.0	10.5	.60	.43
Computational fluency	7.6	3.5	.51	.28
Word identification	100.8	9.9	.47	.49
Sight word fluency	103.1	11.9	.43	.40
Word reading	100.2	11.7	.48	.47
Passage comprehension	96.0	8.8	.49	.58
Cognitive/Behavioral				
Verbal IQ	46.8	9.6	.49	.79
Nonverbal IQ	49.4	10.6	.47	.82
Language	91.5	12.5	.46	.55
Concept formation	92.3	13.3	.50	.51
Long-term memory	92.7	14.1	.33	.39
Working memory	93.7	15.6	.40	.40
Short-term memory	96.7	14.2	.41	.38
Processing speed	97.0	15.5	.42	.31
Attentive behavior	1.4	12.4	.47	.45

Note. All means except math problem solving–nondefinition, word problem solving, computational fluency, verbal and nonverbal IQ, and attentive behavior are based on age-normed standard scores ($M = 100$, $SD = 15$). The math-problem-solving (nondefinition) mean is based on grade-normed standard scores where the mean performance among Grade 3 students tested in the spring is 185. The word-problem-solving mean is based on raw scores with possible scores from 0 to 33. The computational fluency mean is based on raw problems per minute with a maximum possible score of 19.4. The verbal and nonverbal IQ means are based on age-normed T-scores ($M = 50$, $SD = 10$). The attentive behavior mean is based on raw scores with possible scores ranging from -27 to 27 (0 represents a teacher's reported rating of "average" attentive behavior). MPS-def = math problem solving–definition. All correlations are significant, $p < .05$.

Table 2

Effect Sizes Based on Different Definitions of Learning Disability (LA-LD, IQ-ACH).

Measure	LA-LD			IQ-ACH			
	NO LD vs.		LA vs. LD	NO LD vs.		IQ vs. ACH	
	LA	LD		IQ	ACH		
LD definition							
Math problem solving–definition	1.36	2.12	0.76	1.87	1.69	−0.18	
IQ	0.90	1.14	0.24	0.25	1.22	0.97	
Achievement							
Math problem solving–nondefinition	0.99	1.36	0.38	0.90	1.21	0.31	
Word problem solving	0.95	1.32	0.37	0.78	1.21	0.43	
Arithmetic	0.80	1.49	0.69	1.11	1.29	0.17	
Computational fluency	0.62	1.05	0.43	1.03	0.72	−0.32	
Word identification	0.75	0.88	0.13	0.63	0.67	0.04	
Sight word fluency	0.78	0.97	0.19	0.82	0.65	−0.17	
Word reading	0.78	1.09	0.31	0.87	0.86	−0.02	
Passage comprehension	0.73	1.10	0.37	0.77	0.98	0.21	
Cognitive/Behavioral							
Verbal IQ	0.74	0.96	0.21	0.19	1.13	0.94	
Nonverbal IQ	0.72	0.92	0.21	0.21	0.95	0.74	
Language	0.68	0.92	0.24	0.49	0.90	0.41	
Concept formation	0.69	1.11	0.42	0.65	0.98	0.33	
Long-term memory	0.54	0.73	0.19	0.45	0.68	0.23	
Working memory	0.55	0.84	0.29	0.31	0.89	0.58	
Short-term memory	0.60	0.71	0.12	0.35	0.72	0.37	
Processing speed	0.68	0.89	0.21	0.80	0.68	−0.12	
Attentive behavior	0.76	0.98	0.22	0.63	0.81	0.18	

Note. Effect sizes are based on raw or standard scores (see Measures section for scoring) and were computed as the difference in-group mean scores divided by the sample ($n = 813$) standard deviations. Positive effect sizes favor the first group (e.g., for LA vs. LD, LA is the first group, and LD is the second group). For LA versus LD and IQ versus ACH comparisons, large effect sizes ($> .70$) are bolded. LA-LD = low achievement versus extremely low achievement; IQ-ACH = IQ-achievement discrepant versus strictly low-achieving LD; NO LD = students do not meet criteria for low achieving or learning disabled; LA = low achieving in math problem solving (25th percentile, > 10 th percentile), LD = learning disabled in math problem solving based on low achievement (10th percentile), IQ = LD based on discrepancy between IQ and mathproblem solving achievement; ACH = LD based on low achievement in math problem solving only.

Table 3

MANOVA and Regression Results (LA-LD Definitional Group Comparisons).

LA-LD groups	Achievement outcomes							
	mps-nd	wps	arith	comp	wi	sw	rd	pc
MANOVA posthoc comparisons								
NO LD vs. LA	*	*	*	*	*	*	*	*
NO LD vs. LD	*	*	*	*	*	*	*	*
LA vs. LD	*	*	*	*			*	*
Canonicals (LA vs. LD)								
sdfc	0.32	0.23	0.63	0.34	−0.17	−0.20	0.25	0.22
<i>r</i>	.59	.53	.78	.56	.22	.22	.39	.48
Regression coefficient ^a T-values								
NO LD vs. LA/LD	−0.80	0.47	−2.20*	1.01	−0.18	−2.41*	−2.01*	−1.09
LA vs. LD	0.14	1.28	−2.37*	−0.18	1.65	0.54	0.04	−0.18

LA-LD groups	Cognitive/behavioral outcomes								
	viq	nviq	lang	cf	ltm	wm	stm	ps	att
MANOVA posthoc comparisons									
NO LD vs. LA	*	*	*	*	*	*	*	*	*
NO LD vs. LD	*	*	*	*	*	*	*	*	*
LA vs. LD				*		*			
Canonicals (LA vs. LD)									
sdfc	0.09	0.20	0.31	0.59	0.07	0.21	0.04	0.30	0.17
<i>r</i>	.41	.38	.51	.79	.29	.47	.20	.41	.44
Regression coefficient ^a T-values									
NO LD vs. LA/LD	0.10	−0.05	−0.12	−0.76	−1.10	−0.12	0.99	−1.00	−1.20
LA vs. LD	1.18	1.13	0.78	−0.49	−0.20	0.04	1.58	0.59	0.81

Note. LA-LD = low achievement versus extremely low achievement; mps-nd = math problem solving-nondefinition; wps = word problem solving; arith = arithmetic; comp = computational fluency; wi = word identification; sw = sight word fluency; rd = word reading; pc = passage comprehension; NO LD = students do not meet criteria for low achieving or learning disabled; LD = learning disability; LA = low achievement; sdfc = standardized discriminant function coefficient; *r* = canonical correlation; viq = verbal IQ; nviq = nonverbal IQ; lang = language; cf = concept formation; ltm = long-term memory; wm = working memory; stm = short-term memory; ps = processing speed; att = attentive.

^aControlling for math problem solving–definition.

* Significant difference between groups ($p < .05$).

Table 4

MANOVA and Regression Results (IQ-ACH Definitional Group Comparisons).

IQ-ACH groups	Achievement outcomes							
	mps-nd	wps	arith	comp	wi	sw	rd	pc
MANOVA post hoc comparisons								
NO LD vs. IQ	*	*	*	*	*	*	*	*
NO LD vs. ACH	*	*	*	*	*	*	*	*
IQ vs. ACH		*						
Canonicals (IQ vs. ACH)								
sdfc	−0.65	−0.56	−0.40	0.96	−0.08	0.48	0.32	−0.19
<i>r</i>	.47	.48	.17	−.42	.05	−.15	−.02	.24
Regression coefficient ^a T-values								
NO LD vs. IQ/ACH	1.12	1.24	−2.04*	0.45	0.49	−0.5	−1.37	−1.84
IQ vs. ACH	−1.37	−2.31*	−0.89	0.99	0.88	1.82	1.42	1.09

IQ-ACH groups	Cognitive/behavioral outcomes							
	lang	cf	ltm	wm	stm	ps	att	
MANOVA post hoc comparisons								
NO LD vs. IQ	*	*	*	*	*	*	*	
NO LD vs. ACH	*	*	*	*	*	*	*	
IQ vs. ACH				*				
Canonicals (IQ vs. ACH)								
sdfc	0.55	0.20	0.16	0.34	0.56	−0.38	−0.19	
<i>r</i>	.69	.49	.33	.68	.53	−.19	.25	
Regression coefficient ^a T-values								
NO LD vs. IQ/ACH	0.27	−0.39	−0.75	0.39	1.45	−0.17	0.3	
IQ vs. ACH	−0.17	−0.33	0.17	−2.01*	−1.17	0.71	0.1	

Note. IQ-ACH = IQ-achievement-discrepant versus strictly low-achieving learning disability; mps-nd = math problem solving-nondefinition; wps = word problem solving; arith = arithmetic; comp = computational fluency; wi = word identification; sw = sight word fluency; rd = word reading; pc = passage comprehension; NO LD = students do not meet criteria for low achieving or learning disabled; ACH = learning disability based on low achievement in math problem solving only; sdfc = standardized discriminant function coefficient; *r* = canonical correlation; lang = language; cf = concept formation; ltm = long-term memory; wm = working memory; stm = short-term memory; ps = processing speed; att = attentive behavior.

^aControlling for math problem solving-definition and IQ.

* Significant difference between groups ($p < .05$).