Dyscalculia and typical math achievement are associated with individual differences in number specific executive function

Eric D. Wilkey, Courtney Pollack, and Gavin R. Price

Department of Psychology & Human Development, Peabody College, Vanderbilt University, 230 Appleton Place, Nashville, TN, 37203

Corresponding Author

Gavin R. Price

Email: gavin.price@vanderbilt.edu

Abstract:

Deficits in numerical magnitude perception characterize the mathematics learning disability developmental dyscalculia (DD), but recent studies suggest the relation stems from inhibitory control demands from incongruent visual cues in the nonsymbolic number comparison task. This study investigated the relation among magnitude perception during differing congruency conditions, executive function, and mathematics achievement measured longitudinally in children (n = 448) from ages 4 to 13. This relation was investigated across achievement groups and as it related to mathematics across the full range of achievement. Only performance on incongruent trials related to achievement. Findings indicate that executive function in a numerical context, beyond magnitude perception or executive function in a non-numerical context, relates to DD and mathematics across a wide range of achievement.

Mathematical thinking pervades nearly all aspects of modern life, from personal accounting to understanding important information about one's health. Accordingly, individuals with poor mathematical skills are less likely to graduate high school, go to college, have steady employment (Bynner & Parsons, 2006; Rivera-Batiz, 1992), and are at a higher physical and mental health risk (Bynner & Parsons, 2006; Duncan et al., 2007; Hibbard et al., 2007). The development of mathematical skills can be affected by a range of factors including education, home environment, and reading ability. However, a substantial body of research indicates that individual differences in the cognitive system used to perceive and manipulate numerical magnitudes, often labeled the Approximate Number System (ANS) (Feigenson, Dehaene, & Spelke, 2004), play a foundational role in mathematics development (Chen & Li, 2014; Schneider et al., 2017; Schwenk et al., 2017). Further, an estimated 3-6% of the population is affected by the specific mathematics learning disability developmental dyscalculia (DD) (Shalev, Auerbach, Manor, & Gross-Tsur, 2000; Szűcs & Goswami, 2013). Individuals with DD display difficulties with fundamental aspects of numerical processing from very early ages and continue to struggle with math, even when given the same schooling opportunities as their peers. However, the nature of these numerical deficits and their relation to the abilities of typically developing populations remains poorly understood.

The ANS, Mathematics Achievement, and Dyscalculia

The most commonly used behavioral measure of ANS function is the nonsymbolic number comparison task. In this task, participants judge which of two groups of objects, such as dots or squares, is more numerous. Higher accuracy rates and faster response times are thought to indicate higher acuity and enhanced efficiency of the ANS (Inglis & Gilmore, 2014). There is considerable support for a relation between efficiency of the ANS and mathematics achievement, both as a marker for DD (for reviews, see Iuculano, 2016; Szkudlarek & Brannon, 2017), and across the full range of mathematics achievement (for meta-analyses, see Chen & Li, 2014; Schneider et al., 2017).

Accordingly, the dominant theory regarding a core deficit in DD proposes an impairment of the ANS, in part because individuals with DD have been shown to perform more poorly in tasks designed to measure the ANS, such as the nonsymbolic number comparison task (Mazzocco, Feigenson, & Halberda, 2011; Mejias, Mussolin, Rousselle, Grégoire, & Noël, 2012). Further, neuroimaging research suggests that individuals with DD have atypical structure and function of proposed neural substrates of the ANS, such as the intraparietal sulcus (Ashkenazi, Black, Abrams, Hoeft, & Menon, 2013; Dinkel, Willmes, Krinzinger, Konrad, & Koten, 2013; Kaufmann et al., 2009; Mussolin et al., 2010; Price, Holloway, Räsänen, Vesterinen, & Ansari, 2007; Rosenberg-Lee et al., 2015; Rotzer et al., 2008; Rykhlevskaia, Uddin, Kondos, & Menon, 2009). Given this evidence, many researchers suggest that deficits in symbolic number processing, arithmetic fluency, and higher order mathematical thinking stem from a core deficit in the ANS (Butterworth et al., 2011; Iuculano, Tang, Hall, & Butterworth, 2008; Wilson & Dehaene, 2007).

Though there is some consensus that the ANS is atypical in individuals with DD, there is much disagreement as to the true mechanistic nature of this deficit (Szűcs & Goswami, 2013), its causal role in DD (Mazzocco & Räsänen, 2013), and whether the deficit is isolated to the ANS or may be concomitant with deficits in symbolic representation of number or issues related to executive functions (Fias, Menon, & Szűcs, 2013; Rousselle & Noël, 2007; Szűcs, Devine, Soltesz, Nobes, & Gabriel, 2013). It should be further stated that the developmental relation between the ANS and the acquisition of symbolic number faculty is both important and not well understood. It is important in that mathematics is inherently symbolic, and further, most symbolic number tasks have a significantly stronger relation to math achievement than nonsymbolic tasks (De Smedt, Noël, Gilmore, & Ansari, 2013; Fazio, Bailey, Thompson, & Siegler, 2014; Geary et al., 2018; Holloway & Ansari, 2009; Schneider et al., 2017). Therefore, the importance of the ANS for math development may depend on its relation to the acquisition of symbolic number (Reynvoet & Sasanguie, 2016; vanMarle et al., 2018) or their continued relation throughout development (Leibovich & Ansari, 2016), but remains a matter of considerable debate.

Adding to this complication, individual differences in ANS acuity consistently correlate with mathematics across the full range of achievement (Halberda, Mazzocco, & Feigenson, 2008; Keller & Libertus, 2015; Schneider et al., 2017), suggesting the relation is not isolated to group differences that identify severe mathematics deficits, but rather extends broadly across achievement levels. As a result, it remains unclear whether DD represents a qualitatively distinct subgroup with distinct cognitive deficits or is the lowest extreme of a continuous distribution. This distinction is important for developing appropriate intervention strategies to remediate low mathematics skills (Butterworth & Kovas, 2013; Henik, Rubinsten, & Ashkenazi, 2011). For example, if individuals with DD are identified as suffering from a specific impairment of magnitude processing that is qualitatively distinct in its mechanistic origin from their TD peers, it would suggest that remediation should target the training of this uniquely impaired mechanism.

Nonsymbolic Number Comparison as a Measure of the ANS?

One problem undermining the link between ANS function and mathematics development is the reliance on nonsymbolic number comparison as a measure of ANS acuity. Conventionally, nonsymbolic number comparison performance has been interpreted as a measure of ANS function (De Smedt et al., 2013). However, recent research suggests that the task may be measuring more than ANS acuity alone. Specifically, several studies have shown that nonsymbolic number comparison is highly influenced by the visual parameters of task stimuli (Gebuis & Reynvoet, 2011, 2012; Leibovich & Henik, 2013; Szűcs, Nobes, Devine, Gabriel, & Gebuis, 2013). For example, Szűcs et al. (2013) showed that congruency effects have a large impact on the ratio-based internal Weber fraction, or *w*, a common metric of measuring ANS acuity. Further, the impact was even greater for children than in adults, leading them to suggest the visual parameter confound could also be complicated by an interaction with development. In general, visual properties such as surface area and object size covary with numerosity. If these properties are not controlled when creating stimuli, participants can rely on non-numerical cues to select the more numerous array. Thus, to ensure participants employ a strategy focused on numerosity, stimuli are designed such that, in some trials, the more numerous dot set has a greater surface area or dot size (congruent trials), and in other trials a lesser surface area or dot size (incongruent trials) (e.g. Dehaene, Izard, & Piazza, 2005).

Recent studies suggest that performance on incongruent trials may drive the relation between nonsymbolic number comparison and mathematics performance (Bugden & Ansari, 2016; Clayton, Gilmore, & Inglis, 2015; Cragg, Keeble, Richardson, Roome, & Gilmore, 2017; Fuhs & McNeil, 2013; Gilmore et al., 2013; Keller & Libertus, 2015). For example, in a study comparing nonsymbolic number comparison performance in children with DD versus typically developing (TD) peers, Bugden and Ansari (2015) found that children with DD only differed on incongruent trials. A follow-up analysis showed that children's visuo-spatial working memory predicted ANS acuity on incongruent trials, indicating that visuo-spatial working memory may be an important cognitive process utilized for extraction of numerosity in the presence of other visually salient information. Similarly, studies by Gilmore et al. (2013) and Fuhs and McNeil (2013) found that only performance on incongruent trials of the nonsymbolic number comparison task was related to mathematics performance across a wide range of mathematics achievement in primary school and preschoolers respectively. To explain this specific relation, the authors of those studies suggest that incongruent, non-numerical visual cues in the comparison task require participants to inhibit their visually-based response before making a quantity-based judgment, thus engaging inhibitory control mechanisms. Accordingly, both Gilmore et al. and Fuchs and McNeil posit that inhibitory control and selective attention demands of incongruent trials, rather than ANS acuity, drive the relation between nonsymbolic comparison performance and mathematics. Indeed, after controlling for inhibitory control, the relation between mathematics performance and nonsymbolic comparison was no longer statistically significant in both studies.

The ANS and Executive Function

Still, the contribution of executive function to the relation between nonsymbolic number comparison and mathematics performance remains unclear. In contrast to Gilmore et al. (2013) and Fuhs and McNeil (2013), both Keller and Libertus (2015) and Gilmore et al. (2015) found that the relation between accuracy in the number comparison task and mathematics persisted when controlling for inhibitory control, which suggests the relation between number comparison performance and mathematics is not fully accounted for by domain-general inhibitory control. Starr, DeWind, and Brannon (2017) compared the relation between mathematics achievement and the influence of numerical acuity as distinct from the influence of non-numerical visual parameters on nonsymbolic number comparison performance while also measuring inhibitory control in a non-numerical task (i.e. day/night in children and flanker in adults) in a 4- and 6-

year-old sample and a sample of adults. Their results indicated that numerical acuity correlated with higher math scores in the 6-year-old sample while non-numerical bias and inhibitory control did not, which, in agreement with the two previous studies, suggests that numerical discrimination relates to mathematics achievement. However, Starr et al.'s measure of non-numerical bias is a regression term that accounts for the influence of visual parameters on participants' behavior, which is somewhat distinct from performance on trials where visual information is incongruent with numerosity and would more directly address the notion of a number-specific executive function. Further, it should be noted that all five of these studies focused on inhibitory control in a TD sample, while Bugden and Ansari's (2015) findings related performance on incongruent trials of the nonsymbolic comparison task to group differences between DD and TD children. In addition to the group differences versus individual differences distinction between studies, Bugden et al. investigated the role of visuo-spatial working memory as opposed to inhibitory control.

While dominant models indicate that executive function can be divided into the broad categories of working memory/updating, inhibitory control, and attention shifting (Bull & Scerif, 2001; Miyake et al., 2000), most prior studies on nonsymbolic comparison and mathematics achievement have controlled for only one aspect of executive function, either working memory or inhibitory control. As a result, the more fine-grained mechanistic relations between executive function deficits and ANS deficits have been difficult to determine. To address these issues, the current study focuses on two outstanding questions regarding the relation among the ANS, executive function, and mathematics achievement in typically and atypically developing individuals in order to provide more information about the specific mechanisms at play during nonsymbolic number comparison.

First, what are the mechanisms underlying the relation between performance on incongruent trials of the nonsymbolic comparison task and mathematics achievement as compared to congruent trials? Previous studies have framed the correlation between nonsymbolic comparison performance and mathematics achievement as attributable to *either* individual differences in the ANS *or* executive function. An additional possibility is that incongruent trials on the nonsymbolic number comparison task require an *interaction* of executive function and the ANS, or in other words, a *number-specific executive function*. Rather than the relation between number comparison and performance and math achievement depending on neurocognitive mechanisms associated with numerical magnitude processing or executive function independently, a deficit could originate from the biological interplay of these two mechanisms. Successfully answering an incongruent trial requires selective attention to the discrete quantity of each dot set while ignoring other salient yet irrelevant stimulus dimensions. Consistent with this suggestion, experimental studies have demonstrated a distinction between executive function related to numerical and non-numerical content. In a study of DD adults, individuals with DD had difficulty recruiting attention to numerical information but not non-numerical information under heightened cognitive load (Ashkenazi, Rubinsten, & Henik, 2009). In children, Bull and Scerif (2001) demonstrated that inhibitory control and working memory of numerical information accounts for significant variance in individual differences of mathematics ability beyond similar non-numerical measures of executive function. Therefore, to appropriately account for the possibility of an interaction between executive function and the ANS, executive function must be measured in both non-numerical and numerical contexts.

Second, is the relation among executive function, nonsymbolic number comparison, and mathematics achievement a specific facet of atypical development, comprising a characteristic of

DD that sets the disorder qualitatively apart from typical developmental trajectories, or is the relation a characteristic of a broad range of typical mathematics skill development? Previous research appears to suggest that measurements of the ANS correlate with mathematics across the full range of mathematics achievement (Schneider et al., 2017). At the same time, studies suggest that the ANS of individuals with DD is neurobiologically atypical and functions differently than that of their TD peers (Mazzocco et al., 2011; Price et al., 2007). Distinguishing between these alternatives may provide meaningful implications for intervention strategies.

The Current Study

To address the questions above, the current study investigates the relations among ANS function, executive function, and DD by examining performance on the nonsymbolic comparison task, separately for congruent and incongruent trials, while controlling for multiple aspects of executive function. Importantly, executive function here is measured in a non-numerical context. To build directly on previous work, we take a similar approach as Mazzocco et al. (2011). We first compare performance in the nonsymbolic comparison task across multiple mathematics achievement groups (DD, low achieving, and typically achieving) defined through multiple years of consistent achievement, including the first three years of school entry. Second, we consider the relation between performance on the nonsymbolic comparison task and mathematics achievement more broadly through a regression analysis with a large sample that includes the full range of mathematics achievement. In the first analysis, if DD is characterized by a distinct core deficit of the ANS, performance on both congruent and incongruent trials of the task should distinguish among achievement groups. If, on the other hand, DD is characterized by deficits specific to executive function, performance on only the incongruent trials of the nonsymbolic comparison task should account for achievement group differences, but not after controlling for measures of non-numerical executive function. However, if impaired number-specific executive function underlies DD, we would expect group differences between the DD group and the other achievement groups on incongruent trials, but not congruent trials, after controlling for non-numerical, domain-general executive functioning. Similarly, in the second analysis, if number-specific executive function is related to individual differences in mathematics achievement across a wide range of achievement, not only a distinction between DD and the other achievement groups, performance on incongruent trials should predict mathematics achievement beyond what can be accounted for by congruent trials and multiple components of non-numerical executive function.

Method

Participants

The current sample was drawn from a study of students who participated in an earlier longitudinal study of early mathematical skills (Pre-K to 1st grade) (Hofer, Lipsey, Dong, & Farran, 2013). The analytic sample for the original study included 771 children. In the follow-up study, we were able to locate 628 students attending public school in the 2013-14 year in the same district as they attended in Pre-K (16 had withdrawn from the study in 1st grade and were not contacted for further participation, 29 had moved out of the state, 53 had moved out of the district, and 45 were not located despite all efforts). Of those 628, we obtained parental consent and assessed 517 children in the 2013-2014 school year, 506 children in the 2014-2015 school year, and 503 children in the 2015-2016 school year. 497 children were assessed at all three time points in middle school. English language learners (*n* = 43) were excluded because non-native language of mathematics instruction could lead to low mathematics achievement for reasons other than the cognitive factors investigated in the current study.

Our final sample comprised 448 students for whom we had measures of mathematics achievement from 2 of the 3 early time points (spring of Preschool, Kindergarten, and 1st grade) and from 2 of the 3 later time points (5th, 6th, and 7th grades), reading achievement measured at the end of Kindergarten, inhibitory control and task switching measured at 6th or 7th grade, and working memory measured at 5th or 6th grade. This represents a loss of 26 students due to missing data for any of these measures from the full middle school follow-up sample (n = 517), or 5.0%, and only complete cases given the above criteria are analyzed. Methods for resolving differences in measurement year are described below in the description of each measure.

The final sample was 56.5 % female, 9.6% white, 87.1% black, 0.7% Hispanic, 1.1% Middle Eastern, 0.2% Asian or Pacific Islander, and 1.3% other races (no further distinction of race available). Of the 448 students who should have been in 6th grade in the 2014-15 school year if they had not been retained or promoted early, 78 (17.4%) were still in 5th grade and 1 (0.2%) had been promoted to 7th grade. Students were located in 76 schools in the first year of the follow-up study (5th grade), including 31 elementary schools, 27 middle schools, 11 charter schools, and 7 Innovation Cluster schools (i.e., schools that had been targeted for additional resources to boost achievement). Family income level was inferred on the basis of whether participants qualified for free or reduced lunches (i.e., family income less than 1.85 times the U.S. Federal income poverty guideline). In the current sample, 88.6% of participants qualified for free and reduced lunch, 10.3% did not, and 1.1% individuals were missing economic status data. Pre-K through 1st grade and 5th through 7th grade waves of data collection were used to define mathematics achievement groups. Nonsymbolic comparison performance was utilized from 6th grade because concurrent measures of working memory and executive function were available for children in that year. Mean age at the end of pre-K, the first data point, was 5.1 years (*SD* = 0.3, range = 4.5-6.4). See Supplementary Table S1 for full descriptive statistics.

Achievement Groups

Individuals were placed in achievement groups if their mathematics achievement scores were consistently in the designated achievement range at two of the three early assessments (PreK-1st grade) AND two of the three later assessments (5th-7th grades). Given these criteria, 222 children fit into consistent achievement groups across early and later assessment periods, thus excluding 226 children respectively from the full sample of 448 whose achievement level varied beyond the defined threshold across time points. Descriptive statistics for the achievement group sample (*n* = 222) are broken down by achievement group in Table 1.

Our first set of analyses asked whether performance on congruent or incongruent trials of nonsymbolic number comparison distinguished children with DD from their low achieving and typically achieving peers. One commonly used threshold for defining DD is performance in the lowest 10th percentile of standardized mathematics achievement tests (Dinkel et al., 2013; Mazzocco et al., 2011). Several studies comparing groups of student achieving in the lowest 10th percentile to those in the 11th – 25th percentiles reveal important qualitative differences in cognitive profiles (Geary, Hoard, Byrd-Craven, & Nugent, 2007; Mazzocco & Myers, 2003), notably indicating that the lowest achievement group had an impairment in nonsymbolic magnitude processing compared to all other achievement groups (Mazzocco et al., 2011). Therefore, in the current study, we assigned participants to three different mathematics achievement groups, dyscalculic individuals (DD: ≤ 10th percentile), low achieving individuals (LA: 10th - 25th percentile), and typically achieving individuals (TA: 25th - 95th percentile). With these grouping criteria, 22 children met the criteria for DD, 12 for LA, and 188 for TA. Only one individual consistently scored > 95th percentile, a commonly used criterion for school placement in gifted and talented programs, and a common threshold for designating high achieving groups in research (e.g. Hoard, Geary, Byrd-Craven, & Nugent, 2008; Mazzocco et al., 2011). This individual was removed from further analysis.

There is a great diversity in definitions and cutoff thresholds for defining DD in prior literature, and accordingly, findings may not hold across different criteria for selecting DD groups. To address this heterogeneity in the literature, group comparison analyses in the current study were replicated with another commonly used threshold for determining DD (achievement < 1.5 *SD* below the population mean) and included in Appendix A. Using this alternative threshold did not alter the results, suggesting the results are not a product of a chosen threshold.

Many previous studies have attempted to isolate the neurocognitive mechanisms of DD by studying a group of individuals with developmental dyscalculia compared to a control group matched on IQ and other cognitive abilities (Landerl, Bevan, & Butterworth, 2004; Mussolin et al., 2009; Rotzer et al., 2008). The current study does not take this approach for two reasons. First, research suggests that defining learning disability groups through discrepancy criteria excludes many individuals with dyscalculia who suffer from comorbid learning disabilities or other developmental issues. Most estimates suggest that 20-40% of individuals with DD also have dyslexia (Shalev, 2004; Willcutt et al., 2013; Wilson et al., 2015) and around 25% also have attention deficits (Landerl, Göbel, & Moll, 2013; Shalev et al., 1995; Shalev, 2004). This suggests that DD is inherently heterogeneous and would better be characterized by a framework whereby individuals are designated as DD through proof of consistent, low mathematics achievement over time with the presence of adequate educational opportunity (Fuchs, Morgan, Young, & Rise, 2003). Therefore, rather than exclude non-discrepant individuals, the current study follows previous literature (Mazzocco et al., 2011) and investigates differences in the ANS while controlling for reading achievement and domain-general executive function. Second, the current study examines the intersection of attention mechanisms and magnitude processing mechanisms. Any attempt to define groups as a function of broader measures of achievement would impede investigation of individual differences in executive function, which is known to correlate with academic achievement.

--INSERT TABLE 1 APPROXIMATELY HERE--

Procedure

All students assented and students' families consented to participate, and the study was approved by the university's Institutional Review Board (IRB). Assessments were conducted by trained members of the research staff. The nonsymbolic number comparison task and executive function tasks were administered during the Spring semester of the students' 6th grade year via tablet computer. Testing for mathematics achievement was completed in a quiet location at the students' school with one-to-one assistance from trained staff during the student's Pre-K, Kindergarten, 1st grade, 5th grade, 6th grade, and 7th grade years. Reading achievement was assessed at the end of Kindergarten.

Cognitive Tasks

Nonsymbolic number comparison. Participants were presented with two sets of dots simultaneously and asked to indicate via button press which set was more numerous (i.e., which set contained more dots). The set on the left side of the screen contained yellow dots and the set on the right side contained blue dots, which corresponded to color-coded left and right buttons. Response sides were fully counterbalanced. Trials consisted of 1200 ms stimulus presentation followed by 1800 ms of fixation (see Figure 1). Seven ratios were presented [0.33 (5 dots vs. 15

dots), 0.5 (5 vs. 10), 0.67 (6 vs. 9), 0.8 (8 vs. 10), 0.86 (12 vs. 14), 0.88 (7 vs. 8), 0.9 (9 vs. 10)]. The number of dots in each stimulus ranged from 5 to 15. Each ratio was presented 10 times for a total of 70 trials, which were preceded by 6 practice trials of the easiest two ratios.

If individuals did not correctly respond to at least 4 of the 6 practice trials, practice trials were repeated up to two times. If participants did not answer 4 out 6 correctly on any practice run, they did not proceed to the experimental trials. Ratios, stimulus presentation times, and order of presentation were modeled after Odic, Hock and Halberda (2014). To control for the possibility that participants might utilize a strategy based on visual cues rather than number of dots, the following visual properties of dot sets were varied using a modified version of the MATLAB code recommended by Gebuis & Reynvoet (2011): convex hull (area extended by a stimulus), total surface area (aggregate value of dot surfaces), average dot diameter, total circumference, and density (convex hull divided by total surface area). In approximately one quarter of the trials (22 of 70) all four visual properties were congruent with greater numerosity (i.e. the greater number of dots had a greater convex hull, surface area, etc.). In another quarter of the trials (18 of 70), all four visual properties were incongruent with greater numerosity. In the remaining trials, visual properties were mixed congruent and incongruent.

--INSERT FIGURE 1 APPROXIMATELY HERE—

Analyses of task effects include all trials. Analyses directly addressing the research questions include trials that were either fully congruent (22 trials) or incongruent (18 trials) on all five visual parameters. Mixed congruency trials were excluded. Congruent vs. incongruent trials per ratio are not perfectly balanced in trial numbers, but the average ratio for each is nearly identical (average ratio congruent = 0.733, average ratio for incongruent = 0.744)(for further details, see supplementary Table S2). Performance was calculated as mean number of items correct and as a weber fraction (Halberda et al., 2008) to facilitate comparison with previously published research. However, the model implementing Levenberg–Marquardt least squares fit used to calculate weber fractions did not provide a sufficient fit with the fewer number of trials available within congruency conditions (as indicated by whether the model predicted a significant amount of variance, $p < .05$). Further, a growing body of literature suggests that mean accuracy is strongly correlated with and possibly more reliable than ratio dependent metrics such as the weber fraction (Gilmore, Attridge, & Inglis, 2011; Inglis & Gilmore, 2014), which is true even in the case of congruency comparisons (Szűcs et al., 2013). Therefore, in the current study, mean accuracy percentages were used instead of weber fractions to index performance on each of our number comparison tasks.

Working memory. The backward Corsi block-tapping test (Corsi, 1972) provided a measure of visuo-spatial working memory. In this computerized task, children first viewed squares that lit up in a sequence on the screen, and then the students were asked to tap the squares in the reverse order in which they lit up. The task consisted of 16 total possible trials, including two practice trials. The student was given 2 attempts to correctly repeat the reverse sequence per sequence length, increasing in span from 2 to 8 across the task. If the student correctly answered at least 1 of the 2 attempts correctly, the student then proceeded on to the longer (more difficult) sequence. The score of interest was the highest span with a correctly repeated sequence. For some children without 6th grade Corsi spans, 22 children of n = 448, 5th grade spans are utilized. For details, see Appendix B.

Inhibitory control and task switching. The Hearts and Flowers task (Wright & Diamond, 2014) was used as measure of students' task switching and inhibitory control. In this task, the child was first presented with a heart on either side of the screen and instructed to press the button that corresponds to the side of the screen with the heart. This first block comprised 12 trials. In the second block of trials (also 12 trials), the child was presented with flowers and asked to press the button that is opposite the side of the flower. In the third set of trials, the child was randomly presented with either a heart or a flower and asked to follow the rule that corresponds to hearts and flowers respectively. The third block comprised 48 trials. To index executive function we used mean accuracy from the third, mixed-condition block of trials, and as such, this single measure captures both task switching and inhibitory control (Diamond, 2014). One child was not assessed at 6^{th} grade for Hearts and Flowers, but a score from 7^{th} grade was available. The same z-score method described above was utilized to create a score for this child and z-scores were utilized for all subsequent analyses.

Academic Achievement

Reading achievement: Woodcock Johnson III (WCJ-III) – Letter-Word Identification. The WCJ-III (Woodcock, McGrew, & Mather, 2001) is a standard assessment of a range of skills, designed to be used with people ages 2 to 90+. The letter-word identification subtest assesses children's letter and sight word identification ability with the correct pronunciation. Items include identifying and pronouncing letters and words presented to the child (e.g. "A" or "dog"). Age-normed standard scores were calculated as an early measure of reading achievement measured at the end of Kindergarten and then converted to percentile ranks.

Mathematics achievement. Woodcock Johnson-III Quantitative Concepts and Applied problems subtests were used as measures of mathematics achievement during the early school years (Pre-K-1^{st} grade) and KeyMath-3 subscales of Numeration, Algebra, and Geometry were used for the middle school time points (5^{th}-7^{th} grade). Standard scores from each measure were converted to percentile rank scores based on the nationally normed mean and standard deviations of the sample utilized for each respective standardized assessment. Percentile rank scores were utilized for (1) achievement group creations based on percentile rank threshold in the first analysis and (2) the principal outcome variable of interest in our multi-level regression analysis.

WCJ-III – Quantitative Concepts and Applied Problems. Quantitative Concepts and Applied problems subtests were administered at the end of each school year during Pre-K, Kindergarten, and 1st grade. Individually-administered, Quantitative Concepts has two parts and assesses students' knowledge of mathematical concepts, symbols, and vocabulary, including numbers, shapes, and sequences; it measures aspects of quantitative mathematics knowledge and recognition of patterns in a series of numbers. The Applied Problems subtest is an untimed verbal and picture-based measure of a student's ability to analyze and solve mathematics problems, beginning with the application of basic number concepts. At each early time point, age-normed standard scores were calculated for each subtest and averaged together to create a composite measure of mathematics competence representing a broad range of mathematics skills. These scores were subsequently converted to percentile ranks.

KeyMath-3. The KeyMath-3 Diagnostic Assessment (Connolly, 2007) is a comprehensive, norm-referenced measure of essential mathematical concepts and skills. It was administered at the end of each school year during 5th, 6th, and 7th grades. We used three subscales out of the five subscales in the Basic Concepts area. (1) Numeration: The Numeration subtest measures an individual's understanding of whole and rational numbers. It covers topics such as identifying, representing, comparing, and rounding one-, two-, and three-digit numbers as well as fractions, decimal values, and percentages. It also covers advanced numeration concepts such as exponents, scientific notation, and square roots. (2) Algebra: The Algebra subtest measures an individual's understanding of pre-algebraic and algebraic concepts. It covers topics such as sorting, classifying, and ordering by a variety of attributes; recognizing and describing patterns and functions; working with number sentences, operational properties, variables, expressions, equations, proportions, and functions; and representing mathematical relations. (3) Geometry: The Geometry subtest measures an individual's ability to analyze, describe, compare, and classify two-and three-dimensional shapes. It also covers topics such as spatial relations and reasoning, coordinates, symmetry, and geometric modeling. Scale scores in the KeyMath-3 are age-normed to reflect population means of 10 (SD = 3) for each subtest. We averaged scale scores from the three subscales into a composite measure (KM Composite) as in previous analyses involving the current sample (Price & Wilkey, 2017; Rittle-Johnson, Fyfe, Hofer, & Farran, 2017). This score was then converted to a percentile rank to compose mathematics achievement groups across measures of mathematics achievement in the early grades (PreK-1st grade) and late measures of mathematics achievement (5th grade to 7th grade).

The relation between KeyMath-3 scores and predictor variables was non-linear based on visual inspection of scatter plots, so when conducting analyses that assumed a linear relation (e.g. bivariate correlation, partial correlation, or regression), models were fit using a transformed outcome (i.e., cubed root) of KeyMath-3 percentile rank. A detailed exploration of the untransformed achievement scores' relation to predictor variables is detailed in Appendix C.

Analysis

To investigate group differences among DD, LA, and TA groups on nonsymbolic comparison on both congruent and incongruent trials, we conducted a two-way (3 x 2), mixed effects ANOVA with achievement group as a between-subject factor, congruency condition of nonsymbolic comparison as a within-subjects factor, and accuracy rate on the nonsymbolic comparison task at 6th grade as the dependent variable. Levene's tests were run for each ANOVA to analyze violations of homogeneity of variance that often results from unequal sample sizes. When violated, Welch's adjusted F was used for the ANOVA and noted in the results. One-way post-hoc *t*-tests were conducted to examine simple main effects and pairwise differences where appropriate. Bonferroni-corrected *p*-values are reported to correct for multiple comparisons for all subsequent analyses and to ensure tests were robust against violations of homogeneity of variances between groups. Effect sizes are reported as Hedge's *g*, which accounts for unequal group *n*'s by weighting the pooled standard deviation according to group size, $\frac{M_1 - M_2}{SD_{pooled,\ weighted}}$. Because clustering of students within schools did not account for a significant proportion of variation in 6th grade nonsymbolic number comparison accuracy ($\hat{\rho}$ = .009, *p* = .74), a multi-level modeling approach to account for the clustering of students within schools was not needed.

The second set of analyses used random-effects multi-level models to predict 6th grade mathematics achievement from concurrent experimental measures. This analysis examined whether individual differences in nonsymbolic number comparison performance related to standardized mathematics achievement across a wide range of achievement. Specifically, we examined whether 6th graders' accuracy on nonsymbolic number comparison for incongruent and congruent trials predicted concurrent mathematics achievement for the full sample of students (*n* = 448), and whether the relation changed when controlling for early reading achievement and domain-general executive functioning.

Results

Task Effects

Nonsymbolic comparison task performance profiles were consistent with previous findings (e.g., Lyons, Nuerk, & Ansari, 2015), showing a significant effect of ratio on mean accuracy for all trials [$F(6, 447) = 1255.22$, $p < .001$, partial $\eta^2 = 0.737$], and within congruency conditions [$F(6, 447) = 339.01$, $p < .001$, partial $\eta^2 = 0.431$ for congruent trials; $F(6, 447) = 401.17$, $p < .001$, partial $\eta^2 = 0.473$ for incongruent trials]. Further, both mean accuracy and weber fraction were correlated with mathematics achievement at 6th grade (mean accuracy Pearson $r(446) = .191$, $p < .001$, 95% CI [.100, .278]; weber fraction Pearson $r(446) = -.244$, $p < .001$) , 95% CI [-.329, -.155], which is in line with a recent meta-analysis reporting an average correlation of $r = .241$ ($k = 195$) between nonsymbolic comparison and a broad range of mathematics achievement measures across multiple age groups (Schneider et al., 2017). Mean accuracy and weber fractions were highly correlated (Pearson $r(446) = -.919$, $p < .001$) , 95% CI [-.932, -.903].

Achievement Group Comparison Results

Results of the two-way ANOVA indicated a main effect of achievement group [$F(2, 219) = 6.694$, $p = .002$, partial $\eta^2 = 0.058$], a main effect of congruency [$F(1, 219) = 27.570$, $p < .001$, partial $\eta^2 = 0.112$] whereby individuals were more accurate on congruent trials, and an interaction [$F(2, 219) = 4.816$, $p = .009$, partial $\eta^2 = 0.042$]. To characterize the main effect of achievement group, we conducted between-subjects *t*-tests comparing accuracy on the combined congruent and incongruent trials. Accuracy rate was 6.7 points [95% CI: 2.6 – 10.9] lower for the DD group than the LA group [$t(32) = -3.293$, Bonferroni adjusted ($\alpha/3$) $p = .003$, unadjusted $p < .001$, Hedge's $g = 1.182$] and 4.3 points [95% CI: 2.0 – 6.5] lower for the DD group than the TA group [$t(208) = -3.761$, Bonferroni adjusted ($\alpha/3$) $p = .002$, unadjusted $p < .001$, Hedge's $g = 0.847$]. There was no significant difference between the LA and TA groups [$t(198) = 1.619$, Bonferroni adjusted ($\alpha/3$) $p = .161$, unadjusted $p = .053$, Hedge's $g = 0.482$].

The effect of congruency. Pairwise comparisons were conducted to characterize the simple effect of congruency within achievement groups. There was an effect of congruency in the DD and TA groups whereby, on average, the DD group accuracy rate was 25.2 points [95% CI: 16.6 – 33.8] lower for incongruent compared to congruent trials [t(21) = 6.076, Bonferroni adjusted (α/3) $p < .001$, unadjusted $p < .001$, Hedge's $g = 2.203$] and the accuracy rate for the group was 10.7 points [95% CI: 7.6 – 13.8] lower for incongruent compared to congruent trials [t(21) = 6.795, Bonferroni adjusted (α/3) $p < .001$, unadjusted $p < .001$, Hedge's $g = 0.844$]. However, there was no effect of congruency in the LA group [t(11) = 0.716, Bonferroni adjusted (α/3) $p = .732$, unadjusted $p = .244$, Hedge's $g = 0.359$] (see Figure 2 and Table 1 for means).

The effect of achievement group. To characterize the simple effects of achievement group, one-way ANOVAs were conducted within congruency conditions, followed by pairwise comparisons of achievement groups. Results from the ANOVA on accuracy for congruent trials showed no effect of achievement group [$F(2, 219) = .476$, $p = .622$, $\eta^2 = 0.004$] (Figure 2). Levene's test of equality of variances showed no significant differences in variance across groups for mean accuracy of congruent trials (Levene's statistic = .383, $p = .682$).

In contrast, results from the ANOVA on incongruent trials showed a significant effect of achievement group on accuracy [Welch's $F(2, 21.45) = 8.345$, $p = .002$, $\eta^2 = 0.070$]. Levene's test indicated significant differences in variance across groups for mean accuracy of incongruent trials (Levene's statistic = 4.317, $p = .014$), however variance only differed between groups by a factor of 2.56 at most, so Welch's adjusted F was used for the ANOVA. After adjusting for multiple comparisons, post-hoc tests of incongruent trials indicated that accuracy rate for the DD group accuracy rate was 17.3 points [95% CI: 5.2 – 29.4] lower than the LA group [$t(32) = -2.916$, Bonferroni adjusted (α/3) $p = .002$, unadjusted $p = .005$, Hedge's $g = 1.046$] and DD accuracy rate was 12.1 points [95% CI: 5.9 – 18.3] lower than the TA group [$t(208) = -3.862$, Bonferroni adjusted (α/3) $p < .001$, unadjusted $p < .001$, Hedge's $g = 0.870$]. There was and no difference between LA and TA groups [$t(198) = 1.197$, Bonferroni adjusted ($\alpha/3$) $p = .350$, unadjusted $p = .117$, Hedge's $g = 0.356$, mean difference = 5.2 points, 95% CI: -3.3 – 13.7].

To further investigate achievement group differences after controlling for domain-general factors, analyses were repeated as a one-way ANCOVA with the covariates of max span achieved on the backward Corsi, mean accuracy during mixed trials of the Hearts and Flowers task, age at time of testing, and percentile rank on the WCJ-III letter-word identification at the end of Kindergarten. After controlling for these factors, there was still a significant effect of achievement group for accuracy on incongruent trials [$F(2, 215) = 4.658$, $p = .010$, partial $\eta^2 = 0.042$]. After adjusting for multiple comparisons, covariate adjusted means were 16.6 points [95% CI: 3.3 – 29.9] lower for the DD than the LA group [Bonferroni adjusted ($\alpha/3$) $p = .015$, unadjusted $p = .005$, Hedge's $g = 0.823$] and 10.0 points [95% CI: 1.0 – 21.0] lower for the DD group than the TA group [Bonferroni adjusted ($\alpha/3$) $p = .045$, unadjusted $p = .002$, Hedge's $g = 0.823$]. There was no significant difference between the LA and TA groups [Bonferroni adjusted ($\alpha/3$) $p = .231$, unadjusted $p = .077$, Hedge's $g = 0.585$]. These results replicate the pattern observed in the ANOVA.

--INSERT FIGURE 2 APPROXIMATELY HERE—

In sum, all ANOVAs and ANCOVAs conducted show the same pattern of results whereby: (1) no group differences are observed for congruent trials of the comparison task, (2) the DD group performs significantly below LA and TA groups on incongruent trials even when controlling for other cognitive factors and early reading achievement, and (3) no group differences are present between LA and TA groups on incongruent trials.

Full Range of Achievement Results

For descriptive statistics of the full sample, see Table 2. For bivariate correlations among measures, see Supplementary Table S3. Of note is a moderate, negative bivariate correlation

--INSERT TABLE 2 APPROXIMATELY HERE--

between accuracy rates for congruent and incongruent trials ($r(446) = -.447$, $p < .001$, 95% CI [-.518, -.369])(see supplementary Figure S5 for scatterplot). To investigate potential differences among subtests of the KeyMath-3 and their correlations with performance in the nonsymbolic number comparison task, Pearson-*r* values were converted to *z* values and then compared with a two-tailed *z*-test. Results indicated there were no significant differences among any correlations according to KeyMath-3 subtests [all *p*'s > .435] and no further analyses by subtest were conducted (see supplementary Table S4 for details).

Multi-level regression model predicting mathematics achievement. Multi-level modeling accounts for the clustering of students within schools, as approximately 23% of the variation in 6th grade mathematics achievement was due to school membership ($\hat{\rho} = .225$, $p < .0001$). Equation (1) illustrates the modeling approach, in which $MATH_{ij}$ represents 6th grade mathematics achievement for each student *i* in school *j*. The predictors $INCON_{ij}$ and CON_{ij} represent student-level accuracy on nonsymbolic number comparison for incongruent and congruent trials, respectively; HAF_{ij} represents student-level standardized scores on the Hearts and Flowers task; $CORSI_{ij}$ represents student-level standardized backward Corsi max span scores; $READ_{ij}$ represents student-level age-normed standard scores on the letter-word ID test; and $\boldsymbol{X}_{ij}$ represents a vector of potential student-level covariates, such as gender or age at testing. Due to non-linearity in the relation between mathematics scores and the predictors, models were fit using a transformed outcome (i.e., cubed root).

$$\sqrt[3]{MATH_{ij}} = \beta_0 + \beta_1 INCON_{ij} + \beta_2 CON_{ij} + \beta_3 HAF_{ij} + \beta_4 CORSI_{ij} + \beta_5 READ_{ij} + \beta_6 \boldsymbol{X}_{ij} + (e_{ij} + u_j) \qquad (1)$$

The bivariate correlations of the transformed achievement variable are presented in Figure 3 with a plot of nonsymbolic number comparison performance by congruence on achievement.

--INSERT FIGURE 3 APPROXIMATELY HERE--

Table 3 presents parameter estimates, standard errors, significance levels, random effects, and goodness-of-fit statistics for a taxonomy of fitted models describing the relation between mathematics achievement and nonsymbolic number comparison, domain-general executive functioning, early reading achievement, and age at testing in 6th grade. The first model (i.e., M1) displays the grand mean of 6th grade mathematics achievement, across all students and schools, and the intra-class correlation ($\hat{\rho}$ = .225, p < .0001) that motivates the multi-level modeling approach. Model M2 shows the relations between accuracy on congruent and incongruent conditions of the nonsymbolic number comparison task and transformed 6th grade mathematics achievement. There is a statistically significant relation between accuracy on incongruent nonsymbolic number comparison and transformed 6th grade mathematics achievement (z = 4.88, p < .0001), but accuracy on congruent trials is not a statistically significant predictor of mathematics achievement (z = 1.16, p = .25). Accordingly, accuracy on congruent trials was excluded from subsequent models.

Subsequent models (M3-M5) show that the relation between accuracy on incongruent trials of the nonsymbolic number comparison task and transformed 6th grade mathematics achievement persists after controlling for additional predictors of mathematics achievement. Model M3 shows the relation between accuracy on incongruent nonsymbolic number comparison trials and transformed mathematics achievement, controlling for domain-general executive functioning. Hearts and Flowers and backward Corsi performance have a statistically significant relation with mathematics achievement ($z = 7.71$, $p < .0001$ and $z = 7.12$, $p < .0001$, respectively), controlling for nonsymbolic number comparison. Parameter estimates and statistical significance of relations remain stable when controlling for reading performance in Kindergarten (see Table 3, M4) and age of mathematics testing in 6^{th} grade (see Table 3, M5), though the magnitudes decrease slightly. Additional models were fit testing demographic variables (e.g., gender) and interaction terms among the nonsymbolic comparison and executive function predictors, however, none were statistically significant (*p*'s ranged from .06 to .98). Further, we conducted a sensitivity analysis to examine whether students with DD may be driving the relationship between performance on incongruent trials and mathematics achievement. To do so, we refit model M5 without the DD subgroup (n = 22). Results were unchanged. Taken together, the analysis suggests that student performance on incongruent trials of nonsymbolic number comparison is predictive of concurrent mathematics achievement, above and beyond non-numerical, domain-general executive functioning, early reading achievement, and age at testing in 6^{th} grade. For detailed explanation of the model fit, see Appendix B.

--INSERT TABLE 3 APPROXIMATELY HERE—

Discussion

The current study investigated the relation among ANS function, executive function, and mathematics achievement by examining performance on the nonsymbolic comparison task, separately for congruent and incongruent trials, while controlling for multiple components of executive function measured in non-numerical contexts. We investigated this relation first as it relates to group differences among DD, LA, and TA students and then as a factor related to mathematics across a full range of achievement. Results indicated that an interaction of the ANS and executive function mechanisms, beyond either mechanism alone, represents a deficit specific to DD and is also related to mathematics across a full range of mathematics achievement. Together, the current findings suggest that a focus on ANS alone is insufficient to explain the relation between basic number processing and mathematics outcomes. Therefore, we suggest that our results point to a need to reframe existing models of the relation between number processing and mathematics competence to include the relation between executive function mechanisms and magnitude processing, and to move beyond single mechanism explanations more generally.

Achievement Group Comparison

In the first analysis, we compared accuracy rates in the nonsymbolic comparison task across three mathematics achievement levels (i.e. DD, LA, and TA) defined through six years of consistent achievement, including the first three years of school entry (Pre-K-1st grade) and three later years of entry to middle school (5th-7th grade). Our results showed that accuracy on incongruent trials, and not congruent trials, was significantly lower for DD (defined at two different thresholds) compared to LA and TA groups, even after controlling for early reading achievement, visuo-spatial working, inhibitory control, and task shifting. LA and TA groups, on the other hand, did not differ from one another, thus supporting the hypothesis that an impairment in the interaction between executive function and the ANS is characteristic of individuals with DD.

Explanations of the link between ANS and mathematics achievement that involve a dynamic interaction between the ANS and executive function have considerable support from a large body of research linking low mathematics performance with various executive function impairments. These include associations between low mathematics achievement and inhibitory control (Blair & Razza, 2007; Espy et al., 2004; Szűcs et al., 2013), spatial processing (Rourke & Conway, 1997), verbal and visuospatial working memory (Bull & Lee, 2014; Bull & Scerif, 2001; Geary, 2004; Lee & Bull, 2016; Szűcs et al., 2013), set shifting (Willcutt et al., 2013), sustained visual attention (Anobile, Stievano, & Burr, 2013), and inattentive behaviors (Fias et al., 2013; Shalev et al., 1995). Further, DD has a high rate of comorbidity with attention-deficit/hyperactivity disorder (Czamara et al., 2013). Though the link is often made between general measures of executive function and mathematics achievement, there is evidence that the relation is specific to measures of executive function involving numerically relevant information. For example, Siegel and Ryan (1989) found that individuals with DD have impairments of working memory related to processing numerical information and not language. Experimental studies have also demonstrated a distinction between executive function to numerical and non-numerical content. Ashkenazi et al. (2009) found that individuals with DD had more difficulty recruiting attention to numerical information but not non-numerical information under heightened cognitive load compared to TD peers. This array of findings has led some to suggest that DD may involve a domain-specific executive function problem (e.g. Bull & Scerif, 2001). In other words, individuals with DD may not have a generally impaired ANS system, but rather have difficulty working with numerical magnitudes under additional executive function demands. Results from the current study showing mathematics achievement group differences in nonsymbolic comparison performance only during incongruent trials, after controlling for non-numerical executive function, lend further support to this hypothesis. Whether this deficit is driven by a failure to upregulate numerical information above competing information as attention shifting would require, or perhaps a failure to disengage attention from non-numerical information by inhibiting interference from irrelevant stimulus dimensions remains an open empirical question. As the DD group's average performance during incongruent trials is around chance, little can be inferred with about strategy during these trials.

The current study results contrast with some previous studies using an alternative method for controlling visual parameters of dot stimuli which have not found an effect of congruency on response behaviors (Odic et al., 2014; Odic, Libertus, Feigenson, & Halberda, 2013). However, in those studies, the effect of congruency may be confounded by the fact that degree of visual congruency (and incongruency) is linearly related to trial ratio. This means that in difficult ratio trials, which capture the most variance related to individual differences in ANS acuity, each dot set is very similar in terms of surface area, thus decreasing the likelihood of finding a congruency effect. Although this method may be appropriate for measurement of general ANS acuity, the effects of congruency are difficult to separate from the effects of numerical ratio, since the two are linked so tightly. The current study uses a method of controlling congruency that is more balanced across ratios and controls for a greater number of stimulus properties beyond dot size and surface area (for a detailed discussion, see Clayton et al., 2015). Therefore, the effects of congruency and ANS function are more clearly disentangled in the current study.

One unexpected result from the first, group-wise analysis is that DD and TA groups showed congruency effects, as expected, but LA children did not. Despite this lack of a congruency effect in the current findings for this achievement group, we caution against any strong interpretation of this result. There is a trend in the expected direction for each of the LA children groupings (10^{th} percentile and 6.7^{th} percentiles cutoffs), in which children are more accurate on congruent trials than incongruent trials. Despite the lack of a significant effect, the effect sizes are relatively large (Hedge's g = 0.359 and Hedge's g = 0.71) and mean differences are 6 accuracy points and 10 accuracy points for each sample respectively. It is likely that the absence of a statistically significant congruency effect for LA children is due to high variance in accuracy on incongruent trials and a lack of power for this comparison.

Full Range of Achievement

In the second analysis, we examined whether 6th graders' accuracy on nonsymbolic number comparison for incongruent and congruent trials predicted concurrent mathematics achievement for the full sample of students, and whether the relation changed when controlling for early reading achievement and non-numerical, domain-general executive functioning. The sample for this analysis included a wide range of mathematics achievement levels that included all participants from the first analysis and participants in the broader study that did not consistently achieve in the same level year-to-year. Similar to the logic of the first analysis, if number-specific executive function is related to individual differences in mathematics achievement across a wide range of achievement, performance on incongruent trials should predict mathematics achievement beyond what can be accounted for by congruent trials and early reading achievement, visuo-spatial working, inhibitory control, and task shifting. Indeed, results showed that accuracy on incongruent trials predicted concurrent mathematics achievement even after controlling for early reading achievement, visuo-spatial working, inhibitory control, and task shifting, thus supporting the hypothesis that number-specific executive function relates to individual differences in mathematics achievement across a wide range of achievement levels. Further, the relation remained unchanged when we excluded individuals with DD from the regression. These findings build on previous research that has shown other number-specific measures of executive function relate to mathematics achievement in typically developing and high achieving groups. For example, Dark and Benbow (Dark & Benbow, 1994) found that working memory tasks with numerical stimuli were more closely related to mathematical precocity than non-numerical stimuli across a range of tasks in adults. Similarly, studies of children have demonstrated that inhibitory control and working memory of numerical information accounts for significant variance in individual differences of mathematics ability and early numeracy beyond similar non-numerical measures of executive function (Bull & Scerif, 2001; Merkley, Thompson, & Scerif, 2016).

Interestingly, bivariate correlations indicated that children with high accuracy on incongruent trials tended to have low accuracy on congruent trials (and vice versa), even though congruent trials were not related to mathematics achievement. This may be important for two reasons. First, if only incongruent trials are related to mathematics achievement, researchers may be tempted to design measures consisting exclusively of incongruent trials. However, this inverse relation may indicate that incongruent trials are inherently related to congruent trials such that removing congruent trials would change the nature of the task demands for incongruent trials. Second, speculation about inhibitory control has dominated the conversation about the cognitive mechanisms underlying the difference between incongruent trials and congruent trials of the nonsymbolic comparison task (Cragg et al., 2017; Gilmore et al., 2015). While inhibitory control may be a factor, the inverse correlation between congruency conditions may indicate that some individuals are unable to switch between strategies that capitalize on visual cues during congruent trials and ignore these cues otherwise. In addition to working memory and inhibitory control, task shifting may contribute to differences in performance between incongruent and congruent trials. Third, this inverse correlation is somewhat consistent with a developmental account recently suggested by Piazza et al. (2018), whereby development and education both correlate with an increased ability to filter our irrelevant cues in incongruent number comparison trials, similar to those in the current study. In contrast, performance on congruent trials dropped or remained the same with increased education and age, suggesting there was not a generalized increase in acuity of number perception. Piazza et al.'s developmental findings suggests that better performers on incongruent trials may not benefit as much from congruent visual cues, which may explain this inverse correlation.

In the current study, accuracy on congruent trials was unrelated to mathematics achievement, either as a factor distinguishing between achievement groups or as a predictor of mathematics achievement. This was true even before controlling for other academic or cognitive factors. Since current theory suggests engagement of the ANS for successful completion of congruent and incongruent trials, we expected a relation, albeit weaker, between mathematics achievement and accuracy rate on congruent trials. However, neither analysis showed a statistically significant relationship between performance on congruent trials and mathematics achievement. Further, the magnitude of this relationship in both analyses was close to zero, showing no trend in the expected direction. This calls into question whether ANS function alone, not measured under high executive function demands, is an important factor related to DD and mathematics achievement more generally. Studies showing no relation between nonsymbolic number comparison performance and math achievement after controlling for executive function have argued this point. For example, in a large sample of TD children, Szűcs et al. (2014) found that after controlling for other executive function measures such as dot matrices, visuo-spatial working memory, and the trail-making task, nonsymbolic comparison did not significantly relate to mathematics achievement. Interestingly, in that study, sustained visual attention was the best correlate of ANS acuity, which may further indicate that attention mechanisms and ANS mechanisms are integrally related.

Previous neuroimaging research has shown that congruent and incongruent trials of the nonsymbolic number comparison task recruit different neural mechanisms, with incongruent trials recruiting large portions of the fronto-parietal attention network (Leibovich, Vogel, Henik, & Ansari, 2016; Wilkey, Barone, Mazzocco, Vogel, & Price, 2017). Recruitment of additional neurocognitive mechanisms during incongruent trials may be an integral component of the previously assumed direct relation between ANS and mathematics achievement in studies of mathematics learning disability, but also across the full range of achievement. Supporting this interpretation, recent neuroimaging evidence from Wilkey & Price (in press) shows that individual differences in neural activity of inferior frontal brain regions related to the numerical congruency effect in the nonsymbolic comparison task related to mathematics achievement in a typically developing sample of 3rd and 4th grade children. This relation held even after controlling for neural activity in a Flanker task and domain general cognitive factors. In contrast, individual differences in the ratio effect (a neural metric of numerical acuity) did not relate to mathematics, including activity in expected posterior parietal regions. This finding underscores the importance of the neurocognitive mechanisms that interact with magnitude processing mechanisms for mathematics competence, and again speak to the need to move beyond a single mechanism explanation of foundational competencies for mathematics development.

Limitations and Future Directions

Several factors should be taken into account when interpreting the results of the current study. First, participants were recruited from an urban public school system and were mostly from low-income households. Low household income often impedes access to high-quality early mathematics experiences (Ramani & Siegler, 2008), so factors driving the relation between nonsymbolic comparison and mathematics achievement may differ across students with differing household incomes. Further, the relation between nonsymbolic comparison and mathematics achievement in low-income samples has been reportedly lower than middle- and high-income samples (Fuhs, Kelley, O'Rear, & Villano, 2016; Fuhs & McNeil, 2013). However, effect sizes of the relation between nonsymbolic comparison and mathematics achievement from the current study are in line with previous meta-analyses (Chen & Li, 2014; Schneider et al., 2017). Additionally, the lack of relation between mathematics achievement and congruent trials, and significant relation between mathematics achievement and incongruent trials has been previously reported in low-income (Fuhs & McNeil, 2013) and middle-to-high income individuals (Keller & Libertus, 2015). Further, Price and Wilkey (2017) showed that the mediating relation among nonsymbolic comparison accuracy rates and mathematics achievement in the same group of children as the current study follows the same patterns as previously reported literature from wider SES samples (Lyons & Beilock, 2011).

Second, alternative explanations of the current results are possible. For example, rather than our hypothesis about domain-specific executive function, the current results could indicate that individuals who utilize an appropriate strategy for incongruent trials, whether consciously or not, are better at mathematics. If framed as a task strategy, then strategy selection does not necessarily equate to number-specific executive function. Another alternative is that individual differences in task performance are based not on cognitive efficiencies, but rather a predisposition to focus on one aspect of the visual stimuli. A deficit of number-specific executive function is different than the failure to utilize it. Prior research has documented that individuals with a tendency to spontaneously focus on exact quantities have higher arithmetic abilities (Batchelor, Inglis, & Gilmore, 2015; Hannula, Lepola, & Lehtinen, 2010). Recently, this line of research has been expanded to incorporate spontaneous orientation to conflicting or irrelevant dimensions of non-numerical magnitude similar to those of the current study (Viarouge et al., 2017). Research on the underlying neurocognitive mechanisms can also help to distill the root of the differences observed in the current results.

A third factor to consider, specifically in regards to the group comparison results, is that choosing to identify a DD group based on consistent, low mathematics achievement over time has both benefits and limitations for interpreting results. In our methods, we make an argument that DD is likely heterogenous in nature and that identifying a "pure dyscalculic" group via the use of an IQ-math achievement discrepancy criteria results in the exclusion of individuals with DD that do not show a discrepancy due to the high comorbidity of other developmental deficits which would affect IQ or another achievement measure such as reading. With this analytic decision comes the limitation that some individuals within the DD group may perform poorly in mathematics testing due to a more globalized cognitive deficit (e.g. IQ) rather than a specific math deficit, and further, that this global deficit was not adequately controlled for when covarying out reading ability and executive function. One solution suggested by Szűcs (2016) may be to focus more on positioning individuals in a multidimensional parametric space that identifies specific cognitive functions related to mathematical performance. The current results suggest that number-specific executive function is likely to be one such cognitive function.

Fourth, the current study makes the case that number-specific executive function may be impaired in DD and also related to mathematics achievement across a wide achievement range. This conclusion is based on the idea that a relation between two variables (i.e. math achievement and performance on incongruent trials of the number comparison task) survives after controlling for individual differences in other cognitive factors (i.e., executive function in a non-numerical context). In this type of analysis, the conclusion is only as strong as the validity and specificity of control variables. In the current study, only two variables are used as control measures for executive function, and therefore, caution is warranted when considering how completely our variables controlled for all aspects of executive function unrelated to number.

Conclusion

In sum, the two sets of analyses presented here suggest that performance on incongruent trials alone relates to the presence of severe mathematics learning deficits as well as individual differences in mathematics across a wide range of achievement, even when excluding DD individuals. Results suggest that number-specific executive function is a unique predictor of mathematics achievement beyond measures that target the ANS or executive function independently. In order to understand how the intersection of these multiple cognitive mechanisms relates to the acquisition of mathematics skills, future studies should move from a domain-specific vs. domain-general approach to experiments that deconstruct this framework. In so doing, future hypotheses can more closely address the integration of cognitive mechanisms required to complete a complex task such as mathematical thought. Further, the current findings do little to explain the relation between nonsymbolic number perception and symbolic number. Understanding their relation may further explain why number-specific executive function relates to symbolic mathematics. This type of investigation may lead to an enhanced understanding of what type of training or remediation of a specific skill set provides the most potential for transfer to improved mathematics achievement more broadly. Given that the current study provides support for an integral relation between a “domain-general” mechanism with a number-specific one, a training that seeks to leverage this intersection should be explored.

Acknowledgements: This research was supported by the Heising-Simons Foundation (#2013-26) and by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A140126 and R305K050157 to Dale Farran. The opinions expressed are those of the authors and do not represent views of the funders. The authors thank Dale Farran, Kelley Durkin, Kerry Hofer, Jessica Ziegler, Kayla Polk, and Dana True for their assistance with data collection and coding as well as the staff, teachers, and children involved in this research.

References

Anobile, G., Stievano, P., & Burr, D. C. (2013). Visual sustained attention and numerosity sensitivity correlate with math achievement in children. *Journal of Experimental Child Psychology*, *116*(2), 380–391. https://doi.org/10.1016/j.jecp.2013.06.006

Ashkenazi, S., Black, J. M., Abrams, D. A., Hoeft, F., & Menon, V. (2013). Neurobiological Underpinnings of Math and Reading Learning Disabilities. *Journal of Learning Disabilities*, *46*(6), 549–569. https://doi.org/10.1177/0022219413483174

Ashkenazi, S., Rubinsten, O., & Henik, A. (2009). Attention, automaticity, and developmental dyscalculia. *Neuropsychology*, *23*(4), 535–540. https://doi.org/10.1037/a0015347

Batchelor, S., Inglis, M., & Gilmore, C. (2015). Spontaneous focusing on numerosity and the arithmetic advantage. *Learning and Instruction*, *40*, 79–88. https://doi.org/10.1016/j.learninstruc.2015.09.005

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Bugden, S., & Ansari, D. (2016). Probing the nature of deficits in the ‘Approximate Number System’ in children with persistent Developmental Dyscalculia. *Developmental Science*, *19*(5), 817–833. https://doi.org/10.1111/desc.12324

Bull, R., & Lee, K. (2014). Executive Functioning and Mathematics Achievement. *Child Development Perspectives*, *8*(1), 36–41. https://doi.org/10.1111/cdep.12059

Bull, R., & Scerif, G. (2001). Executive Functioning as a Predictor of Children’s Mathematics

Ability: Inhibition, Switching, and Working Memory. *Developmental Neuropsychology*, *19*(3), 273–293. https://doi.org/10.1207/S15326942DN1903_3

Butterworth, B., & Kovas, Y. (2013). Understanding Neurocognitive Developmental Disorders Can Improve Education for All. *Science*, *340*(6130), 300–305. https://doi.org/10.1126/science.1231022

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: from brain to education. *Science (New York, N.Y.)*, *332*(6033), 1049–1053. https://doi.org/10.1126/science.1201536

Bynner, J., & Parsons, S. (2006). *Does Numeracy Matter More? National Research and Development Centre for Adult Literacy and Numeracy*. London: The Basic Skills Agency. Retrieved from http://www.nrdc.org.uk/publications_details.asp?ID=16#

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychologica*, *148*, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica*, *161*, 177–184. https://doi.org/10.1016/j.actpsy.2015.09.007

Connolly, A. J. (2007). KeyMath-3 Diagnostic Assessment. San Antonio, TX: Pearson.

Corsi, P. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International*, *34*(02).

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, *162*, 12–26.

https://doi.org/10.1016/j.cognition.2017.01.014

Czamara, D., Tiesler, C. M. T., Kohlb??ck, G., Berdel, D., Hoffmann, B., Bauer, C. P., … Heinrich, J. (2013). Children with ADHD Symptoms Have a Higher Risk for Reading, Spelling and Math Difficulties in the GINIplus and LISAplus Cohort Studies. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0063859

Dark, V. J., & Benbow, C. P. (1994). Type of Stimulus Mediates the Relationship Between Working-Memory Performance and Type of Precocity. *Intelligence*, *19*, 337–357. https://doi.org/10.1016/0160-2896(94)90006-X

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children’s mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

Dehaene, S. (2011). *The number sense: How the mind creates mathematics.* Oxford University Press.

Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*.

Diamond, A. (2014). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dinkel, P. J., Willmes, K., Krinzinger, H., Konrad, K., & Koten, J. W. (2013). Diagnosing developmental dyscalculia on the basis of reliable single case FMRI methods: promises and limitations. *PloS One*, *8*(12), e83722. https://doi.org/10.1371/journal.pone.0083722

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology*, *26*(1), 465–486. https://doi.org/10.1207/s15326942dn2601_6

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*(1), 53–72. https://doi.org/10.1016/j.jecp.2014.01.013

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Fias, W., Menon, V., & Szucs, D. (2013). Multiple components of developmental dyscalculia. *Trends in Neuroscience and Education*, 1–5. https://doi.org/10.1016/j.tine.2013.06.006

Fuchs, D., Morgan, P. L., Young, C. L., & Rise, T. (2003). Responsiveness-to-Intervention : Definitions , Evidence , and Implications for the Learning Disabilities Construct, *18*(3), 157–171.

Fuhs, M. W., Kelley, K., O’Rear, C., & Villano, M. (2016). The Role of Non-Numerical Stimulus Features in Approximate Number System Training in Preschoolers from Low-Income Homes. *Journal of Cognition and Development*, *17*(5), 737–764. https://doi.org/10.1080/15248372.2015.1105228

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Developmental Science*, *16*(1), 136–148. https://doi.org/10.1111/desc.12013

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, *37*(1), 4–15. https://doi.org/10.1177/00222194040370010201

Geary, D. C., Hoard, M. K., Byrd-craven, J., & Nugent, L. (2007). Cognitive Mechanisms Underlying Achievement Deficits in Children With Mathematical Learning Disability, *78*(4), 1343–1359.

Geary, D. C., VanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early Conceptual Understanding of Cardinality Predicts Superior School-Entry Number-System Knowledge. *Psychological Science*, *29*(2), 191–205. https://doi.org/10.1177/0956797617729817

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods*, *43*(4), 981–986. https://doi.org/10.3758/s13428-011-0097-5

Gebuis, T., & Reynvoet, B. (2012). Continuous visual properties explain neural responses to nonsymbolic number. *Psychophysiology*, *49*(11), 1649–1659. https://doi.org/10.1111/j.1469-8986.2012.01461.x

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., … Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PloS One*, *8*(6), e67374. https://doi.org/10.1371/journal.pone.0067374

Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *Quarterly Journal of Experimental Psychology (2006)*, *64*(11), 2099–2109. https://doi.org/10.1080/17470218.2011.574710

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *Zdm*, 1–12. https://doi.org/10.1007/s11858-014-0659-y

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(October), 665–668. https://doi.org/10.1038/nature07246

Hannula, M. M., Lepola, J., & Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *Journal of Experimental Child Psychology*, *107*(4), 394–406. https://doi.org/10.1016/j.jecp.2010.06.004

Henik, A., Rubinsten, O., & Ashkenazi, S. (2011). The “where” and “what” in developmental dyscalculia. *The Clinical Neuropsychologist*, *25*(6), 989–1008. https://doi.org/10.1080/13854046.2011.599820

Hibbard, J. H., Peters, E., Dixon, A., Tusler, M., Peters, E., & Dixon, A. (2007). Consumer Competencies and the Use of Comparative Quality Information: It Isn’t Just about Literacy. *Medical Care Research and Review*, *64*(4), 379–394. https://doi.org/10.1177/1077558707301630

Hoard, M. K., Geary, D. C., Byrd-craven, J., & Nugent, L. (2008). Mathematical Cognition in Intellectually Precocious First Graders. *Developmental Neuropsychology*, *33*(3), 251–276. https://doi.org/10.1080/87565640801982338

Hofer, K. G., Lipsey, M. W., Dong, N., & Farran, D. C. (2013). *Results of the Early Math Project – Scale-Up Cross-Site Results*. Nashville, TN.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147–155. https://doi.org/10.1016/j.actpsy.2013.11.009

Iuculano, T. (2016). *Neurocognitive accounts of developmental dyscalculia and its remediation. Progress in brain research* (1st ed., Vol. 227). Elsevier B.V. https://doi.org/10.1016/bs.pbr.2016.04.024

Iuculano, T., Tang, J., Hall, C. W. B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science*, *11*(5), 669–680. https://doi.org/10.1111/j.1467-7687.2008.00716.x

Kaufmann, L., Vogel, S. E., Starke, M., Kremser, C., Schocke, M., & Wood, G. (2009). Developmental dyscalculia: compensatory mechanisms in left intraparietal regions in response to nonsymbolic magnitudes. *Behavioral and Brain Functions : BBF*, *5*, 35. https://doi.org/10.1186/1744-9081-5-35

Keller, L., & Libertus, M. (2015). Inhibitory control may not explain the link between approximation and math abilities in kindergarteners from middle class families. *Frontiers in Psychology*, *6*(May), 1–11. https://doi.org/10.3389/fpsyg.2015.00685

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9-year-old students. *Cognition*, *93*(2), 99–125. https://doi.org/10.1016/j.cognition.2003.11.004

Landerl, K., Göbel, S. M., & Moll, K. (2013). Core deficit and individual manifestations of developmental dyscalculia (DD): The role of comorbidity. *Trends in Neuroscience and Education*, *2*(2), 38–42. https://doi.org/10.1016/j.tine.2013.06.002

Lee, K., & Bull, R. (2016). Developmental changes in working memory, updating, and math achievement. *Journal of Educational Psychology*, *108*(6), 869–882. https://doi.org/10.1037/edu0000090

Leibovich, T., & Ansari, D. (2016). The symbol-grounding problem in numerical cognition: A review of theory, evidence, and outstanding questions. *Canadian Journal of Experimental Psychology*, *70*(1), 12–23. https://doi.org/10.1037/cep0000070

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*(June), 375. https://doi.org/10.3389/fpsyg.2013.00375

Leibovich, T., Vogel, S. E., Henik, A., & Ansari, D. (2016). Asymmetric Processing of Numerical and Nonnumerical Magnitudes in the Brain: An fMRI Study. *Journal of Cognitive Neuroscience*, *28*(1), 166–176. https://doi.org/10.1162/jocn_a_00887

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–261. https://doi.org/10.1016/j.cognition.2011.07.009

Lyons, I. M., Nuerk, H.-C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology. General*, *144*(5), 1021–1035. https://doi.org/10.1037/xge0000094

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired Acuity of the Approximate Number System Underlies Mathematical Learning Disability (Dyscalculia). *Child Development*, *82*(4), 1224–1237. https://doi.org/10.1111/j.1467-8624.2011.01608.x

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia*. https://doi.org/10.1007/s11881-003-0011-7

Mazzocco, M., & Rasanen, P. (2013). Contributions of longitudinal studies to evolving definitions and knowledge of developmental dyscalculia. *Trends in Neuroscience and Education*. Elsevier. https://doi.org/10.1016/j.tine.2013.05.001

Mejias, S., Mussolin, C., Rousselle, L., Grégoire, J., & Noël, M.-P. (2012). Numerical and nonnumerical estimation in children with and without mathematical learning disabilities. *Child Neuropsychology*, *18*(6), 550–575. https://doi.org/10.1080/09297049.2011.625355

Merkley, R., Thompson, J., & Scerif, G. (2016). Of huge mice and tiny elephants: Exploring the relationship between inhibitory processes and preschool math skills. *Frontiers in Psychology*, *6*(JAN), 1–14. https://doi.org/10.3389/fpsyg.2015.01903

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mussolin, C., De Volder, A., Grandin, C., Schlögel, X., Nassogne, M.-C., & Noël, M.-P. (2010). Neural Correlates of Symbolic Number Comparison in Developmental Dyscalculia. *Journal of Cognitive Neuroscience*, *22*(5), 860–874. https://doi.org/10.1162/jocn.2009.21237

Mussolin, C., De Volder, A., Grandin, C., Schlogel, X., Nassogne, M. C., & Noel, M. P. (2009). Neural Correlates of Symbolic Number Comparison in Developmental Dyscalculia. *Journal of Cognitive Neuroscience*, (Early Access), 1–15.

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology. General*, *143*(1), 255–265. https://doi.org/10.1037/a0030825

Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental Change in the Acuity of Approximate Number and Area Representations. *Developmental Psychology*, *49*(6), 1103–1112. https://doi.org/10.1037/a0029472

Piazza, M., De Feo, V., Panzeri, S., & Dehaene, S. (2018). Learning to focus on number. *Cognition*, *181*(July), 35–45. https://doi.org/10.1016/j.cognition.2018.07.011

Price, G. R. G. R., & Wilkey, E. D. E. D. (2017). Cognitive mechanisms underlying the relation between nonsymbolic and symbolic magnitude processing and their relation to math. *Cognitive Development*, *44*(September), 139–149. https://doi.org/10.1016/j.cogdev.2017.09.003

Price, G. R., Holloway, I. D., Räsänen, P., Vesterinen, M., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Current Biology*, *17*(24), R1042–R1043. https://doi.org/10.1016/j.cub.2007.10.013

Price, G. R., & Wilkey, E. D. (2017). Cognitive mechanisms underlying the relation between nonsymbolic and symbolic magnitude processing and their relation to math. *Cognitive Development*, *44*. https://doi.org/10.1016/j.cogdev.2017.09.003

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*(2), 375–394. https://doi.org/10.1111/j.1467-8624.2007.01131.x

Reynvoet, B., & Sasanguie, D. (2016). The Symbol Grounding Problem Revisited: A Thorough Evaluation of the ANS Mapping Account and the Proposal of an Alternative Account Based on Symbol–Symbol Associations. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01581

Rittle-Johnson, B., Fyfe, E. R., Hofer, K. G., & Farran, D. C. (2017). Early Math Trajectories: Low-Income Children's Mathematics Knowledge From Ages 4 to 11. *Child Development*, *88*(5), 1727–1742. https://doi.org/10.1111/cdev.12662

Rivera-Batiz, F. L. (1992). Quantitative Literacy and the Likelihood of Employment among Young Adults in the United States. *The Journal of Human Resources*, *27*(2), 313–328. Retrieved from http://www.jstor.org/stable/145737

Rosenberg-Lee, M., Ashkenazi, S., Chen, T., Young, C. B., Geary, D. C., & Menon, V. (2015). Brain hyper-connectivity and operation-specific deficits during arithmetic problem solving in children with developmental dyscalculia. *Developmental Science*, *18*(3), 351–372. https://doi.org/10.1111/desc.12216

Rotzer, S., Kucian, K., Martin, E., von Aster, M., Klaver, P., Loenneker, T., … Loenneker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia.

NeuroImage, *39*(1), 417–422. https://doi.org/10.1016/j.neuroimage.2007.08.045

Rourke, B. P., & Conway, J. A. (1997). Disabilities of arithmetic and mathematical reasoning perspectives from neurology and neuropsychology. *Journal of Learning Disabilities*, *30*(1), 34–46.

Rousselle, L., & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, *102*(3), 361–395. https://doi.org/10.1016/j.cognition.2006.01.005

Rykhlevskaia, E., Uddin, L. Q., Kondos, L., & Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: combined evidence from morphometry and tractography. *Frontiers in Human Neuroscience*, *3*(November), 51. https://doi.org/10.3389/neuro.09.051.2009

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, *20*(3), e12372. https://doi.org/10.1111/desc.12372

Schwenk, C., Sasanguie, D., Kuhn, J. T., Kempe, S., Doebler, P., & Holling, H. (2017). (Non-)symbolic magnitude processing in children with mathematical difficulties: a meta-analysis. *Research in Developmental Disabilities*, *64*, 152–167. https://doi.org/10.1016/j.ridd.2017.03.003

Shalev, R. S. (2004). Developmental dyscalculia. *Journal of Child Neurology*, *19*(10), 765–771.

Shalev, R. S., Auerbach, J., & Gross-Tsur, V. (1995). Developmental dyscalculia behavioral and attentional aspects: a research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *36*(7), 1261–1268. https://doi.org/10.1111/j.1469-7610.1995.tb01369.x

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia: prevalence and prognosis. *European Child & Adolescent Psychiatry*, *9*(S2), S58–S64. https://doi.org/10.1007/s007870070009

Siegel, L. S., & Ryan, E. B. (1989). The Development of Working Memory in Normally Achieving and Subtypes of Learning Disabled Children. *Child Development*, *60*(4), 973. https://doi.org/10.2307/1131037

Starr, A., DeWind, N. K., & Brannon, E. M. (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition*, *168*, 222–233. https://doi.org/10.1016/j.cognition.2017.07.004

Szkudlarek, E., & Brannon, E. M. (2017). Does the Approximate Number System Serve as a Foundation for Symbolic Mathematics? *Language Learning and Development*, *13*(2), 171–190. https://doi.org/10.1080/15475441.2016.1263573

Szűcs, D. (2016). Subtypes and comorbidity in mathematical learning disabilities: Multidimensional study of verbal and visual memory processes is key to understanding. In *Progress in Brain Research* (Vol. 227, pp. 277–304). https://doi.org/10.1016/bs.pbr.2016.04.027

Szűcs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, *49*(10), 2674–2688. https://doi.org/10.1016/j.cortex.2013.06.007

Szűcs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2014). Cognitive components of a mathematical processing network in 9-year-old children. *Developmental Science*, *17*(4), 506–524. https://doi.org/10.1111/desc.12144

Szűcs, D., & Goswami, U. (2013). Developmental dyscalculia: Fresh perspectives. *Trends in Neuroscience and Education*, *2*(2), 33–37. https://doi.org/10.1016/j.tine.2013.06.004

Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology*, *4*(July), 444. https://doi.org/10.3389/fpsyg.2013.00444

vanMarle, K., Chu, F. W., Mou, Y., Seok, J. H., Rouder, J., & Geary, D. C. (2018). Attaching meaning to the number words: contributions of the object tracking and approximate number systems. *Developmental Science*, *21*(1), 1–17. https://doi.org/10.1111/desc.12495

Viarouge, A., Courtier, P., Hoppe, M., Melnik, J., Houdé, O., & Borst, G. (2017). Spontaneous orientation towards irrelevant dimensions of magnitude and numerical acuity. *Learning and Instruction*, 1–8. https://doi.org/10.1016/j.learninstruc.2017.09.004

Wilkey, E. D., Barone, J. C., Mazzocco, M. M. M., Vogel, S. E., & Price, G. R. (2017). The effect of visual parameters on neural activation during nonsymbolic number comparison and its relation to math competency. *NeuroImage*, *159*(August), 430–442. https://doi.org/10.1016/j.neuroimage.2017.08.023

Wilkey, E.D., & Price, G. R. (in press). Attention to number: The convergence of numerical magnitude processing, attention, and mathematics in the inferior frontal gyrus. Human Brain Mapping, https://doi.org/ 10.1002/hbm.24422

Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., & Pennington, B. F. (2013). Comorbidity Between Reading Disability and Math Disability. *Journal of Learning Disabilities*, *46*(6), 500–516. https://doi.org/10.1177/0022219413477476

Wilson, A. J., & Dehaene, S. (2007). Human behavior, learning, and the developing brain: Atypical development. In *Number sense and developmental dyscalculia.* (pp. 212–238).

Wilson, A. J., Andrewes, S. G., Struthers, H., Rowe, V. M., Bogdanovic, R., & Waldie, K. E. (2015). Dyscalculia and dyslexia in adults: Cognitive bases of comorbidity. *Learning and Individual Differences*, *37,* 118–132. https://doi.org/10.1016/j.lindif.2014.11.017

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock–Johnson III Tests of Achievement. Itasca, IL: Riverside.

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology*, *5*(MAR), 1–9. https://doi.org/10.3389/fpsyg.2014.00213

Table 1. Descriptive statistics for achievement subgroups.

Achievement Group Sample	**DD** (n = 22, 7 females)			**LA** (n = 12, 6 females)			**TA** (n = 188, 106 females)		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Age (years), Pre-K	5.1	0.5	4.5-6.4	5.0	0.3	4.7-5.5	5.1	0.3	4.5-5.6
Age (years), 6^{th} grade	12.2	0.5	11.4-13.4	12.0	0.3	11.6-12.5	12.0	0.3	11.4-12.6
Nonsymbolic Comparison (accuracy, %)	71.5	5.3	62.9-81.4	78.2	6.4	70.0-87.1	75.8	5.0	58.6-91.4
Nonsymbolic Comparison (Congruent accuracy, %)	78.7	9.1	63.6-90.9	76.9	10.7	54.5-86.4	76.3	11.1	40.9-95.5
Nonsymbolic Comparison (Incongruent accuracy, %)	53.5	13.3	27.8-83.3	70.8	21.3	33.3-94.4	65.7	14.0	33.3-94.4
Nonsymbolic Comparison (weber fraction, *w*)	0.37	0.11	0.21-.65	0.26	0.10	.13-.48	0.27	0.07	0.10-0.56
Backward Corsi* (z-score of max span)	-1.21	1.22	-2.4-0.95	0.03	0.57	-0.75-0.95	0.37	0.85	-2.44-2.65
Hearts and Flowers* (z-score of accuracy, %)	-1.29	0.79	-2.33-0.82	-0.16	0.83	-1.90-1.83	0.40	0.83	-1.90-1.83
Letter-Word Identification (WCJ-III, standard score)	91.4	9.90	75-113	97.4	11.9	73-113	115.1	11.9	85-144

* z-scores presented based on full sample of 448 individuals.
WCJ-III = Woodcock Johnson III. KM-3 = KeyMath-3.

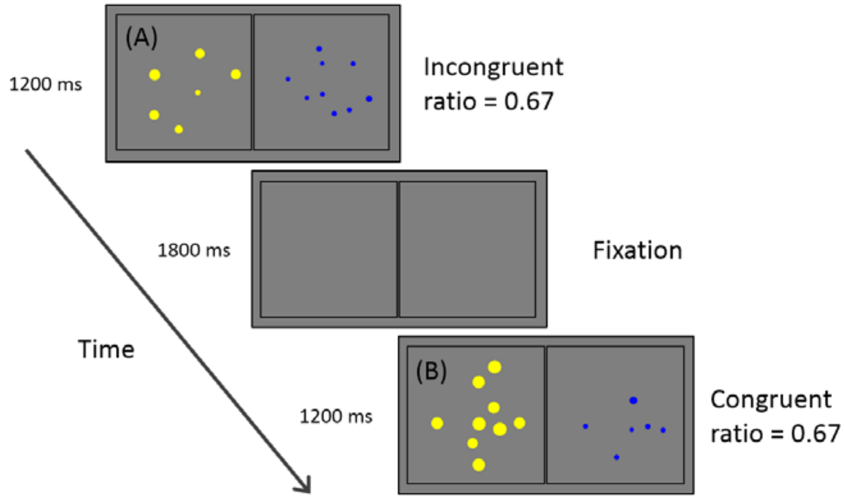

Figure 1. Nonsymbolic numerical magnitude comparison stimuli and paradigm timing. (A) Incongruent trial example of ratio 0.67 (smaller number dot set/larger number dot set, 6/9 = 0.67). (B) Congruent trial example, also of ratio 0.67.

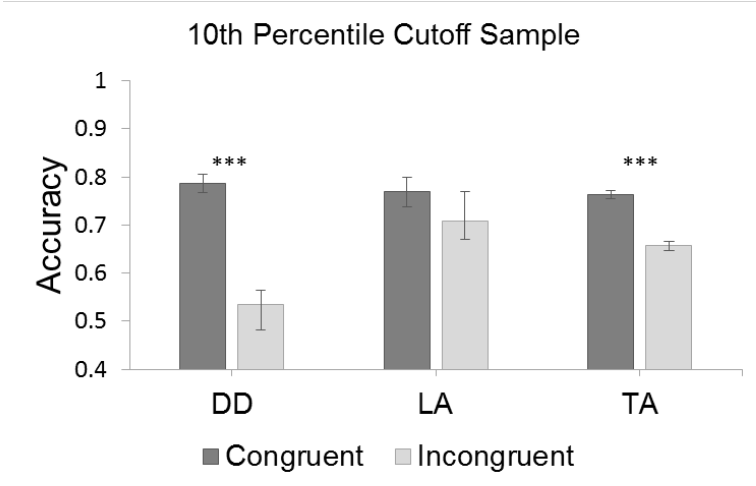

Figure 2. Nonsymbolic number comparison accuracy rates by achievement group. DD = developmental dyscalculia. LA = low achieving. TA = typically achieving. Error bars represent standard errors. P-values are indicated for differences in accuracy between congruent and incongruent trials (*** *p* < .001).

Table 2. Descriptive statistics for experimental and standardized measures for full sample.

	Entire Sample (n = 448, 250 females)		
	Mean	SD	Range
Age (years), 6th grade	12.0	.32	11.4-13.4
Nonsymbolic Comparison (accuracy, %)	74.8	5.48	48.6-91.4
Nonsymbolic Comparison (Congruent accuracy, %)	76.6	11.2	36.4-100
Nonsymbolic Comparison (Incongruent accuracy, %)	63.1	14.5	22.2-94.4
Nonsymbolic Comparison (weber fraction, *w*)	0.29	0.10	0.10-1.42
Backward Corsi * (max span)	4.81	1.22	2-8
Hearts and Flowers* (accuracy, %)	73.4	14.5	35-100
Letter-word ID – WCJ-III (K, percentile rank)	109.7	12.7	73-144
Math Achievement - KM-3 (6th grade, percentile rank)	27.0	23.1	0.5-92.5

* Raw scores reported here for year available. See sections 2.4.2 and 2.4.3 for a detailed description of scores used for analyses. WCJ-III = Woodcock Johnson III. KM-3 = KeyMath-3.

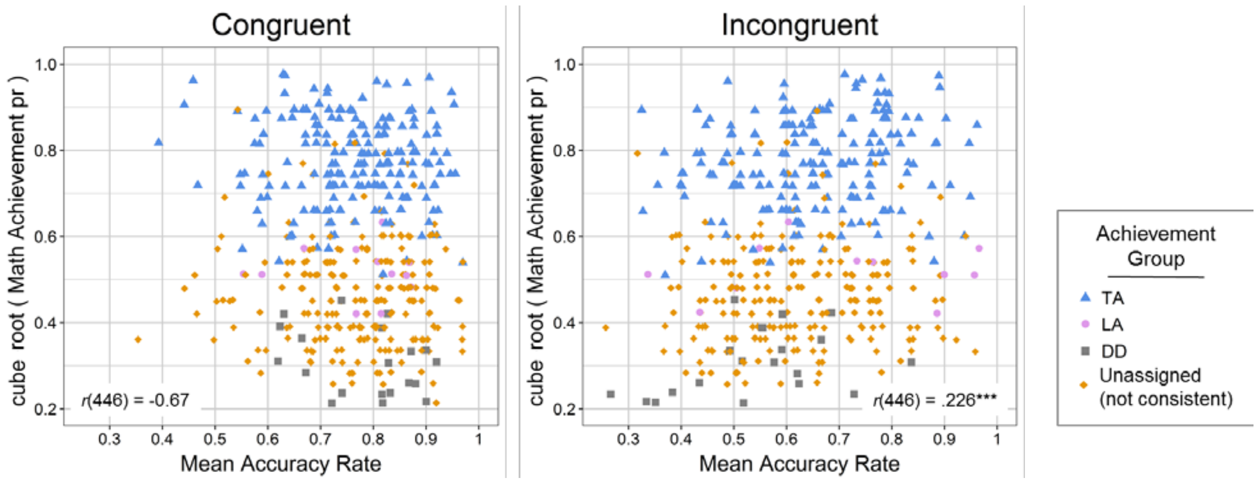

Figure 3. Nonsymbolic number comparison accuracy rates split by (left) congruent and (right) incongruent trials including all individuals from the full sample plotted against $\sqrt[3]{MATH_{ij}}$, the outcome variable of equation 1 below, cube root of the composite math achievement percentile rank. DD = developmental dyscalculia. LA = low achieving. TA = typically achieving. pr = percentile rank. Bivariate correlations of the full sample are presented in the bottom corner of each panel (*** $p < .001$). Orange diamonds represent individuals who did not fit our selection criteria for stable achievement grouping based on Pre-K to 7th grade achievement.

Table 3. Taxonomy of fitted multi-level models describing the relation between the cubed root of 6th grade mathematics achievement and accuracy on nonsymbolic number comparison, separately for incongruent and congruent trials, controlling for working memory, inhibitory control and task switching, reading achievement, and age of testing in 6th grade ($n_{schools}$ = 75; $n_{students}$ = 448).

	6th grade mathematics achievement (cubed root)				
	M1	M2	M3	M4	M5
Intercept	0.562*** (0.015) [.532–.593]	0.286** (0.091) [.107–.465]	0.480*** (0.036) [.410–.551]	0.011 (0.072) [-.131–.152]	-1.084** (0.337) [-1.744–-.424]
Nonsymbolic Comparison, incongruent trials, acc.		0.321*** (0.066) [.192–.449]	0.144** (0.053) [.039–.248]	0.141** (0.051) [.042–.241]	0.126* (0.050) [.027–.224]
Nonsymbolic Comparison, congruent trials, acc.		0.097 (0.083) [-.066–.261]			
Backward Corsi, max span			0.054*** (0.008) [.039–.069]	0.051*** (0.007) [.037–.065]	0.050*** (0.007) [.036–.064]
Hearts and Flowers, mixed trials, acc.			0.060*** (0.008) [.045–.075]	0.053*** (0.007) [.038–.067]	0.051*** (0.007) [.036–.065]
Reading achievement, LWID, end of Kindergarten				0.004*** (0.001) [.003–.006]	0.005*** (0.001) [.004–.007]
Age of KeyMath-3 testing, 6th grade					0.007*** (0.002) [.003–.011]
$\hat{\sigma}_u$	0.094*** (0.014)	0.091*** (0.014)	0.073*** (0.011)	0.057*** (0.010)	0.051*** (0.010)
$\hat{\sigma}_e$	0.174*** (0.006)	0.170*** (0.006)	0.150*** (0.005)	0.144*** (0.005)	0.143*** (0.005)
$\hat{\rho}$	0.225*** (0.057)	0.223*** (0.057)	0.190*** (0.050)	0.134*** (0.043)	0.112** (.040)
Log-likelihood	114.559	126.743	185.816	211.721	217.176

*p < .05, **p < .01, ***p < .001. Acc. = accuracy, LWID = letter-word identification, $\hat{\sigma}_u$ = School-level residual standard deviation, $\hat{\sigma}_e$= Student-level residual standard deviation, $\hat{\rho}$ = Intra-class correlation. Standard errors are in parentheses and 95% confidence intervals are in brackets.

Appendix A. Detailed Results from 6.7th Percentile cutoff sample achievement group analysis.

To make current results more easily comparable to previous literature that used differing cutoff thresholds for determining DD groups, the current study also examined whether there were differences between two commonly used thresholds for determining a dyscalculic sample. This threshold has varied widely across studies, and has likely contributed to disagreement among findings (Mazzocco & Myers, 2003). Another commonly used threshold is mathematics achievement scores 1.5 standard deviations below the nationally normed means, which is equivalent to performance below the 6.7th percentile (Kaufmann et al., 2013; Price et al., 2007; Rotzer et al., 2009). This threshold resulted in the following achievement groupings: DD, $\leq$ 6.7th percentile; LA, 6.7th - 25th percentile; TA, 25th - 95th percentile. Again, individuals were placed in achievement groups if their mathematics achievement scores were consistently in the designated achievement range at two of the three early assessments (PreK-1st grade) AND two of the three later assessments (5th-7th grades). Given these criteria, 221 children fit into consistent achievement groups across early and later assessment periods, 11 children met the criteria for DD, 22 for LA, and the same 188 children were TA. Descriptive statistics in Table A1.

Results

As in the first achievement group sample, there were no differences according to gender distribution percentages of mathematics achievement groups with the 6.7th percentile cutoff grouping (Pearson $\chi^2(2) = 4.045$, $p = .132$, Cramer's V = .132), nor in mathematics achievement ($t(446) = 1.182$, $p = .238$, Cohen's $d = 0.112$) or in nonsymbolic comparison accuracy ($t(446) = 0.780$, $p = .436$, Cohen's $d = 0.074$) at 6th grade, the outcome year of interest for the second set of primary analyses.

Table A1. Descriptive statistics for experimental and standardized measures.

	10th Percentile Cutoff Sample (n = 222, 116 females)			**6.7th Percentile Cutoff Sample** (n = 221, 115 females)			**Entire Sample** (n = 448, 250 females)		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Age (years), Pre-K	5.1	0.3	4.5-6.4	5.1	0.3	4.5-6.4			
Age (years), 6th grade	12.0	0.3	11.4-13.4	12.0	0.3	11.4-13.4	12.0	.32	11.4-13.4
Nonsymbolic Comparison (accuracy, %)	75.5	5.29	58.6-91.4	75.6	5.3	58.6-91.4	74.8	5.48	48.6-91.4
Backward Corsi * (max span)	5.1	1.2	2-8	5.2	1.1	2-8	4.81	1.22	2-8
Hearts and Flowers* (accuracy, %)	76.4	14.4	40-100	76.8	13.9	44-100	73.4	14.5	35-100
Letter-word ID – WCJ-III (K, percentile rank)	111.8	14.1	73-144	111.8	14.2	73-144	109.7	12.7	73-144
Math Achievement- WCJ-III (Pre-K, percentile rank)	51.3	24.9	1.0-95.0	52.4	23.8	1.0-95.0			
Math Achievement- WCJ-III (K, percentile rank)	52.1	24.7	0.0-93.0	52.7	23.8	0.0-93.0			
Math Achievement- WCJ-III (1st grade, percentile rank)	48.1	24.6	0.4-95.5	48.6	24.1	0.4-95.5			
Math Achievement - KM-3 (5th grade, percentile rank)	39.2	23.5	0.5-96.2	39.6	23.1	0.7-96.2			
Math Achievement - KM-3 (6th grade, percentile rank)	42.1	22.7	0.5-92.5	42.4	22.3	1.0-92.5	27.0	23.1	0.5-92.5
Math Achievement - KM-3 (7th grade, percentile rank)	42.6	22.9	0.5-94.1	42.9	22.5	0.5-94.1			

* Raw scores reported here for year available. See sections 2.4.2 and 2.4.3 for a detailed description of scores used for analyses. WCJ-III = Woodcock Johnson III. KM-3 = KeyMath-3.

Detailed Results from 6.7th Percentile cutoff sample achievement group analysis. For the 6.7th percentile cutoff sample, there was an effect of congruency in the DD and TA groups [$t(10) = 3.855$, $p = .003$, Cohen's $d = 1.968$ for DD; $t\,(187) = 6.795$, $p < .001$, Cohen's $d = 0.844$ for TA], but not in the LA group [$t\,(21) = .705$, $p = .068$, Cohen's $d = 0.705$]. The right panel of Figure A2 shows the congruency effect for DD and TA groups in the 6.7 percentile cutoff sample. Levene's test of equality of variances showed no significant differences in variance across groups for mean accuracy of congruent trials or incongruent trials. Results from the ANOVA showed that there was no effect of achievement group on number comparison performance for congruent trials [$F(2, 218) = .389$, $p = .679$., $\eta^2 = 0.003$], but there was a significant effect of achievement group on number comparison performance for incongruent trials [$F(2, 218) = 4.947$, $p = .008$, $\eta^2 = 0.043$]. After adjusting for multiple comparisons, one-tailed post-hoc tests indicated lower accuracy rates for DD than TA children (Bonferroni adjusted $p = .003$, Hedge's $g = 0.997$), lower accuracy rates for DD than LA children (Bonferroni adjusted $p = .011$, Hedge's $g = 0.821$), and no difference between LA and TA groups (Bonferroni adjusted $p = .500$, Hedge's $g = 0.028$).

Results from the ANCOVAs with the covariates of mean accuracy on the Hearts and Flowers mixed trials, max span on the backward Corsi block-tapping test, age at grade 6 testing, and letter-word identification at the end of Kindergarten indicated there was no effect of achievement group on number comparison performance for congruent trials [$F(2, 214) = .208$, $p = .812$, partial $\eta^2 = 0.002$], but there was a significant effect for incongruent trials [$F(2, 214) = 3.356$, $p = .037$, partial $\eta^2 = 0.030$]. After adjusting for multiple comparisons, one-tailed post-hoc tests indicated lower accuracy rates for DD than TA children (Bonferroni adjusted $p = .034$, Hedge's $g = 0.895$ lower accuracy rates for DD than LA (Bonferroni adjusted $p = .017$, Hedge's $g = 0.893$), and no difference between LA and TA groups (Bonferroni adjusted $p = .500$, Hedge's $g = 0.112$). These results replicate the pattern observed in the ANOVA.

The same ANOVA's and ANCOVA's were conducted on groups formed with the 6.7th percentile cutoff threshold for both congruent and incongruent trials and results fit the same pattern as those of the 10th percentile cutoff. In sum, all ANOVA's and ANCOVA's conducted on both the 10th and 6.7th percentile cutoff samples show the same pattern of results whereby: (1) no group differences are observed for congruent trials of the nonsymbolic comparison task, (2) the DD group performs significantly below LA and TA groups on incongruent trials even when controlling for other cognitive factors and early reading achievement, and (3) no group differences are present between LA and TA groups on incongruent trials.

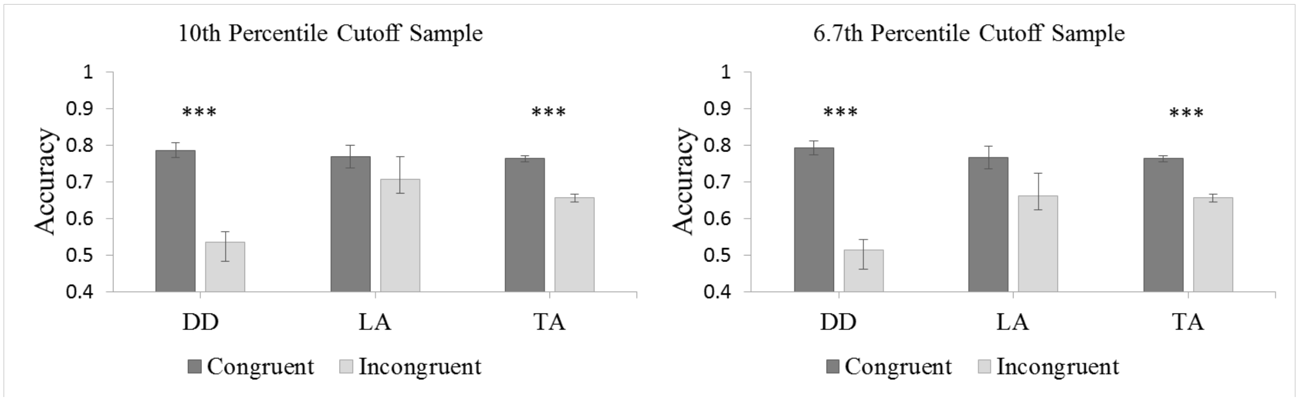

Figure A2. Nonsymbolic number comparison accuracy rates for the sample with developmental dyscalculia (DD) defined as achievement below the 10th percentile (left) and 6.7th percentile (right) split by congruency. LA = low achieving. TA = typically achieving. Error bars represent standard errors. P-values are indicated for differences in accuracy between congruent and incongruent trials (*** p < .001).

Appendix B. Details of missing 6th grade Corsi span scores.

At the 6th grade assessment, 22 children (of n = 448) did not proceed from instruction in the backward Corsi to successful completion of a trial, indicating noncompliance with the task or a failure to understand instructions. Scores on outcome measures and covariates of interest for these children were different, on average, from those children who successfully completed the task (nonsymbolic accuracy $t(446) = 3.728$, $p < .001$, Cohen's $d = 0.794$; Hearts and Flowers mean accuracy $t(446) = 3.508$, $p < .001$, Cohen's $d = 0.716$; 6th grade mathematics achievement $t(446) = 2.587$, $p = .010$, Cohen's $d = 0.613$). Therefore, to avoid nonrandom missing data and include these children in our analyses, backward Corsi max span from the 5th grade was used, where available. To maintain the relative position of children's scores in the 5th grade among other children's 6th grade scores (5th grade mean max span = 4.52, 6th grade mean span = 4.88), both years of backward Corsi max spans were z-scored and 5th grade z-scores of the 22 children were used instead of 6th grade z-scores, which were used for the other 426 children.

Appendix C. Exploration of the model fit.

In order to better interpret the non-linear relation between accuracy on incongruent trials of the nonsymbolic number comparison task and mathematics achievement, we plot this relation in Figure 3. This figure shows the fitted relation between untransformed 6th grade mathematics achievement and nonsymbolic number comparison accuracy on incongruent trials for Model M6, holding Hearts and Flowers accuracy, backward Corsi span, early reading achievement, and age at testing in 6th grade at their sample means. As Figure 3 shows, the magnitude of the relation between accuracy on incongruent trials and mathematics achievement is greater for students with higher accuracy, on average. For example, the estimated difference between students with 30% and 40% accuracy on nonsymbolic number comparison is associated with a difference of 1.0 percentile rank points in 6th grade mathematics achievement, on average. The difference between students with 75% and 85% accuracy on nonsymbolic number comparison is associated with a difference of 1.3 percentile rank points in 6th grade mathematics achievement, on average.

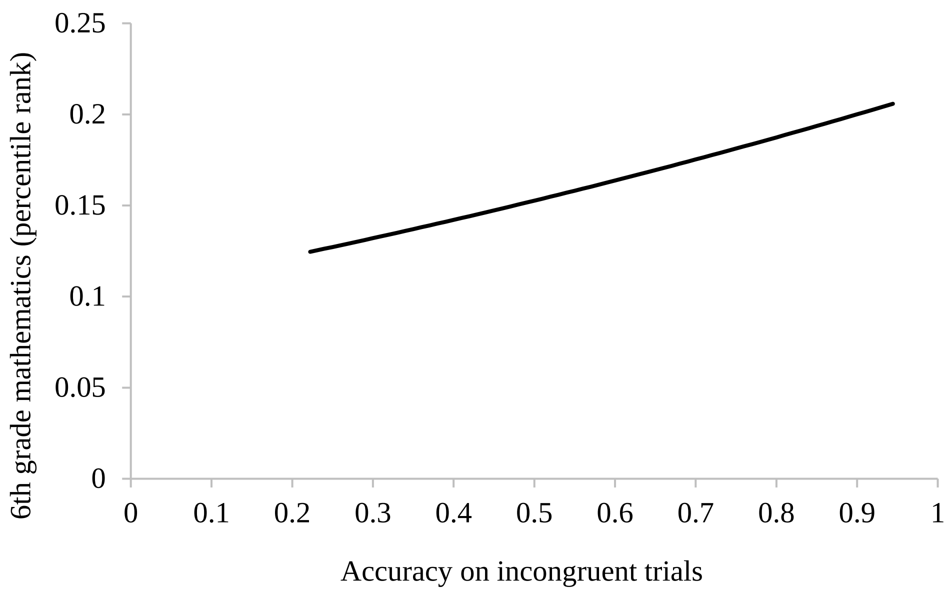

Figure B1. Predicted 6th grade mathematics achievement as a function of accuracy on incongruent trials of nonsymbolic number comparison, for students with average domain-general executive functioning and early reading achievement, and of average age at testing in 6th grade.

Table S1. Descriptive statistics for experimental and standardized measures.

	Achievement Group Sample (n = 222, 116 females)			**Entire Sample** (n = 448, 250 females)		
	Mean	SD	Range	Mean	SD	Range
Age (years), Pre-K	5.1	0.3	4.5-6.4			
Age (years), 6th grade	12.0	0.3	11.4-13.4	12.0	.32	11.4-13.4
Nonsymbolic Comparison (accuracy, %)	75.5	5.29	58.6-91.4	74.8	5.48	48.6-91.4
Nonsymbolic Comparison (Congruent accuracy, %)				76.6	11.2	36.4-100
Nonsymbolic Comparison (Incongruent accuracy, %)				63.1	14.5	22.2-94.4
Nonsymbolic Comparison (weber fraction, *w*)`	0.28	0.08	0.10-0.65	0.29	0.10	0.10-1.42
Backward Corsi * (max span)	5.1	1.2	2-8	4.81	1.22	2-8
Hearts and Flowers* (accuracy, %)	76.4	14.4	40-100	73.4	14.5	35-100
Letter-word ID – WCJ-III (K, percentile rank)	111.8	14.1	73-144	109.7	12.7	73-144
Math Achievement- WCJ-III (Pre-K, percentile rank)	51.3	24.9	1.0-95.0			
Math Achievement- WCJ-III (K, percentile rank)	52.1	24.7	0.0-93.0			
Math Achievement- WCJ-III (1st grade, percentile rank)	48.1	24.6	0.4-95.5			
Math Achievement - KM-3 (5th grade, percentile rank)	39.2	23.5	0.5-96.2			
Math Achievement - KM-3 (6th grade, percentile rank)	42.1	22.7	0.5-92.5	27.0	23.1	0.5-92.5
Math Achievement - KM-3 (7th grade, percentile rank)	42.6	22.9	0.5-94.1			

* Raw scores reported here for year available. See sections 2.4.2 and 2.4.3 for a detailed description of scores used for analyses. WCJ-III = Woodcock Johnson III. KM-3 = KeyMath-3.

Supplementary Table S2. Task details for number comparison tasks of all formats, *n* = 448.

Nonsymbolic Number Comparison Task Details and Accuracy Rates By Ratio					
Ratios	Accuracy (%) all trials (*SD*)	Accuracy (%) CON (*SD*)	Accuracy (%) INC (*SD*)	Trials (CON)	Trials (INC)
0.33 (5 v 15)	0.984	0.977	0.983	2	2
0.5 (5 v 10)	0.958	0.935	0.974	2	2
0.67 (6 v 9)	0.925	0.949	0.844	4	2
0.8 (8 v 10)	0.697	0.743	0.472	6	2
0.86 (12 v 14)	0.640	0.701	0.394	4	2
0.88 (7 v 8)	0.504	0.488	0.500	2	6
0.9 (9 v 10)	0.530	0.502	0.512	2	2

SD= standard deviation. CON = congruent. INC = incongruent. Average ratio for congruent trials was 0.733 and for incongruent trials was 0.744

Table S3. Pearson *r* values for bivariate correlations between measures included in regression model predicting 6th grade mathematics achievement.

Measure (n = 448)	**1**	**2**	**3**	**4**	**5**
1. Nonsymbolic Comparison, congruent trials, acc.					
2. Nonsymbolic Comparison, incongruent trials, acc.	-.447*** [-.518, -.369]				
3. Backward Corsi, max span	-.051 [-.143, .042]	.193*** [.102, .280]			
4. Hearts and Flowers, mixed trials, acc.	.017 [-.076, .109]	.186*** [.095, .274]	.242*** [.153, .327]		
5. Reading achievement, LWID, end of Kindergarten	-.010 [-.102, .083]	.071 [-.022, .162]	.130** [.038, .220]	.172*** [.081, .260]	
6. Mathematics achievement, composite, grade 6	-.067 [-.158, .026]	.226*** [.136, .312]	.396*** [.315, .471]	.411*** [.331, .485]	.412*** [.332, .486]

* $p < .05$, ** $p < .01$, *** $p < .001$. Acc. = accuracy; LWID = letter-word identification (WCJ-III). The mathematics achievement composite score is cube-root transformed as described below. 95% Confidence Intervals in brackets.

Supplementary Table S4. Comparison of nonsymbolic comparison performance and Keymath-3 subtest correlations**.**

		KeyMath-3 Number Subscale: Age-scaled Score	KeyMath-3 Algebra Subscale: Age-scaled Score	KeyMath-3 Geometry Subscale: Age-scaled Score
Nonsymbolic Comparison (accuracy, %)	Pearson Correlation	.151**	.166**	.197**
	Sig. (2-tailed)	.001	.000	.000
	Fischer *r*-to-*z*; z test	Number vs Algebra *p*= .818	Algebra vs Geometry *p* = .631	Geometry vs Number *p* = .478
Nonsymbolic Comparison (Congruent accuracy, %)	Pearson Correlation	-.086	-.052	-.040
	Sig. (2-tailed)	.069	.274	.397
	Fischer *r*-to-*z*; z test	Number vs Algebra *p*= .610	Algebra vs Geometry *p* = .857	Geometry vs Number *p* = .490
Nonsymbolic Comparison (Incongruent accuracy, %)	Pearson Correlation	.196**	.182**	.232**
	Sig. (2-tailed)	.000	.000	.000
	Fischer *r*-to-*z*; z test	Number vs Algebra *p*= .826	Algebra vs Geometry *p* = .435	Geometry vs Number *p* = .576
Nonsymbolic Comparison (weber fraction, w)	Pearson Correlation	-.200**	-.231**	-.209**
	Sig. (2-tailed)	.000	.000	.000
	Fischer *r*-to-*z*; z test	Number vs Algebra *p*= .631	Algebra vs Geometry *p* = .734	Geometry vs Number *p* = .889

* Pearson r values were compared by transforming the r value to a z value and comparing the z values using a two-tailed z test with an alpha of .05.

Figure S5. Scatter plots of nonsymbolic number comparison performance measures and composite math achievement (A, B, C) and a plot of congruent by incongruent accuracy rates (D).