Kindergartners' Fluent Processing of Symbolic Numerical Magnitude is Predicted by Their
Cardinal Knowledge and Implicit Understanding of Arithmetic Two Years Earlier

Alex M. Moore, Kristy vanMarle, and David C. Geary

University of Missouri

Corresponding author:
David C. Geary, Ph.D.
Curators' Professor
Department of Psychological Sciences
Interdisciplinary Neuroscience Program
210 McAlester Hall
University of Missouri
Columbia, MO 65211-2500
E-mail: GearyD@Missouri.edu
Ph: 573-882-6268

Words (abstract, text, references):6,142

Acknowledgments: The study was supported by grants from the University of Missouri Research Board and DRL- 1250359 from the National Science Foundation. We thank Mary Rook for her help in facilitating our assessments of the Title I preschool children. We are also grateful for the cooperation of Columbia Public Schools and especially the children and parents involved in the study. We thank Tim Adams, Melissa Barton, Sarah Beckett, Sam Belvin, Erica Bizub, Kaitlyn Bumberry, Lex Clarkson, Stephen Cobb, Danielle Cooper, Dillon Falk, Lauren Johnson-Hafenschner, Jared Kester, Morgan Kotva, Brad Lance, Kayla Legow, Natalie Miller, Lexi Mok, Molly O'Byrne, Rebecca Peick, Kelly Regan, Nicole Reimer, Laura Roider, Sara Schroeder, Claudia Tran, Hannah Weise, Melissa Willoughby, and Grace Woessner for help with data collection and entry, Mary Hoard and Lara Nugent for help with managing the project, and Jeff Rouder for consultation on some of the analyses.

Abstract

Fluency in first graders' processing of the magnitudes associated with Arabic numerals, collections of objects, and mixtures of objects and numerals predicts current and future mathematics achievement. The quantitative competencies that support the development of fluent processing of magnitude are not fully understood, however. At the beginning and end of preschool ($M = 3$ years, 9 months at first assessment; range 3 years, 3 months to 4 years, 3 months), 112 (51 boys) children completed tasks measuring numeral recognition and comparison, acuity of the approximate number system, and knowledge of counting principles, cardinality, and implicit arithmetic, and completed a magnitude processing task (number sets test) in kindergarten. Use of Bayesian and linear regression techniques revealed that two measures of preschoolers' cardinal knowledge and their competence at implicit arithmetic predicted later fluency of magnitude processing, controlling domain general factors, preliteracy skills, and parental education. The results help to narrow the search for the early foundation of children's emerging competence with symbolic mathematics and provide direction for early interventions.

Keywords: number sets test, preschool, kindergarten, development, mathematical cognition, cardinality, approximate number system

Introduction

Children who start school behind their peers in mathematics are at high risk of staying behind throughout their academic careers and into adulthood (Duncan et al., 2007; Geary, Hoard, Nugent, & Bailey, 2013; Jordan, Kaplan, Ramineni, & Locuniak, 2009; Morgan, Farkas, & Wu, 2009; Ritchie & Bates, 2013; Shalev, Manor, & Gross-Tsur, 2005). Identifying and eventually remediating the core competencies that predict poor school-entry mathematics achievement has the potential to reduce this risk. Indeed, screening tests for early quantitative deficits now exist, and generally focus on children's implicit and explicit understanding of number and their ability to operate (e.g., count) on these representations (Gersten et al., 2012 for a review). Most of these studies however have focused on predictors of mathematics achievement in kindergarten or later, although there is now some evidence that younger children's and even infants' sensitivity to relative magnitude may predict later achievement (e.g., Starr, Libertus, & Brannon, 2013), but this is vigorously debated (below).

We examined the relation between preschoolers' nonsymbolic and symbolic quantitative abilities and their kindergarten performance on a measure of the fluent processing of the magnitudes associated with Arabic numerals. We chose the latter because it captures early number knowledge that prospectively predicts children's risk of mathematical learning disability (MLD) in elementary school (Geary, Hoard, & Bailey, 2009) and individual differences in mathematics achievement more generally (e.g., Cirino, Fuchs, Elias, Powell, & Schumacher, 2015), controlling domain-general abilities (e.g., working memory) and demographic factors. By bridging performance on measures of early quantitative competencies and performance on a number knowledge measure that predicts later mathematics achievement, we help to narrow the

search for early quantitative knowledge that sets the stage for the longer-term development of mathematical competence.

Quantitative foundations

In recent years, there has been a flurry of studies of the relation between preschoolers' quantitative abilities and their mathematics achievement, but no consensus has yet been reached regarding which of these early abilities is foundational (e.g., Bonny & Lourenco, 2013; Chu, vanMarle, & Geary, 2015; Libertus, Halberda, & Feigenson, 2011; Mazzocco, Feigenson, & Halberda, 2011; Star, Libertus, & Brannon, 2013; vanMarle, Chu, Li, & Geary, 2014). Many of these studies have focused on the approximate number system (ANS) that supports infants' and nonhuman animals' ability to discriminate the relative quantities of collections of objects (Feigenson, Dehaene, & Spelke, 2004; Geary, Berch, & Mann Koepke, 2015, for reviews). Acuity of these discriminations is the central developmental and individual difference feature of the ANS. Newborn infants can reliably discriminate the quantities of collections of objects that differ by a 3:1 ratio (Izard, Sann, Spelke, & Streri, 2009), 3:2 at 9 months (e.g., Brannon, Suanda, & Libertus, 2007; Libertus & Brannon, 2010), and up to a ratio of 11:10 in adulthood (Halberda & Feigenson, 2008). Beyond quantity discrimination, the ANS may also allow for an intuitive understanding of arithmetical operations (e.g., Gallistel & Gelman, 1992). The combination of mechanisms for representing the approximate value of collections of objects and for manipulating these representations may provide the scaffolding for preschoolers' emerging understanding of Arabic numerals and the relations among them.

Crucially, there is some evidence that individual differences in ANS acuity contribute to individual differences in children's in mathematical achievement generally, and perhaps to MLD (e.g., Halberda et al., 2008; Bonny & Lourenco, 2013; Libertus et al., 2011; Lourenco, Bonny,

Fernandez, & Rao, 2012; Mazzocco et al., 2011; Piazza et al., 2010). These early findings were promising; however, the results from subsequent studies have been more mixed (Bugden & Ansari, 2011; De Smedt & Gilmore, 2011; De Smedt, Noël, Gilmore, & Ansari, 2013; Lyons, Ansari, & Beilock, 2012; Matejko & Ansari, 2016; Negen & Sarnecka, 2014; Rousselle & Noël, 2007; Sasangui, Defever, Maertens, & Reynoet, 2014). The gist of these latter studies is that young children's general mathematics achievement may be more strongly related to the speed and accuracy with which they can compare or manipulate (e.g., order) the magnitudes associated with Arabic numerals than to performance on measures of ANS acuity (for meta-analyses see Chen & Li, 2014; Fazio, Bailey, Thompson, & Seigler, 2014). However, most of these studies have been conducted with kindergarten or older children, leaving open the possibility that ANS acuity has a time- and content-limited relation to preschoolers' early learning of symbolic mathematics.

Indeed, in their meta-analysis, Fazio et al. (2014) found that the relation between ANS acuity and mathematics achievement was stronger before ($r = .4$) rather than after ($r = .17$) children begin formal instruction. Other studies suggest that the relation between ANS acuity and mathematics achievement may be mediated by children's understanding of the cardinal value of number words (Chu et al., 2015; vanMarle et al., 2014), controlling executive functions, intelligence, preliteracy skills, and parental education. Nevertheless, the unsettled state of the field means that it would be premature to exclude the ANS from the search for the early foundations of children's mathematical development. Accordingly, we included a standard measure of children's ANS acuity (Halberda et al., 2008) and a measure of their ability to engage in nonverbal addition and subtraction that could in theory be supported by the ANS (Gelman, 2006; Levine, Huttenlocher, & Jordan, 1992; McCrink & Spelke, 2016).

At the same time, the search for this foundation must also include preschoolers' emerging knowledge and understanding of number symbols. This necessarily involves learning the count list, "one, two, three...", but also the quantities associated with these number words, that is, their cardinal values (Gelman & Gallistel, 1978; Sarnecka & Gelman, 2004). Although many two-year olds can recite the count list up to "ten", it often takes two more years before they understand the associated cardinal values (Gelman, 2006; Wynn, 1990, 1992). This developmental progression is revealed in the give-a-number task in which children are asked to give an adult a specific number of objects. Most preschoolers initially give exactly 'one' when requested, but give a random number of objects for all other number words. Such a child is described as a "one-knower". Three- to six-months later, children become "two-knowers" and a few months later "three-knowers". Once children understand the quantities represented by 'three' or 'four', they induce the meaning of the rest of the count list to become "cardinal principle-knowers" (CP-knowers); these children give the exact number requested up to the limit of their count list (Carey, 2009; Le Corre & Carey, 2007; Wynn, 1992).

Accordingly, we administered the give-a-number task and a second measure of children's understanding of the cardinal value of number words. We also assessed children's recognition of Arabic numerals, their ability to discriminate the larger of two such numerals, and their understanding of how number words are used during the act of counting (Gelman & Meck, 1983); the latter is the basis for early, symbolic arithmetic (Siegler & Shrager, 1984).

Fluent number processing

The above noted studies of symbolic skills have largely focused on older children's competence in accessing and comparing the magnitudes associated with Arabic numerals (e.g., Bugden & Ansari, 2011; De Smedt & Gilmore, 2011; De Smedt et al., 2013; Lyons et al., ,

2012; Rousselle & Noël, 2007). Bugden and Ansari, for instance, administered several tasks that tapped 1st and 2nd graders' automatic and intentional processing of the magnitudes associated with Arabic numerals and based on their findings concluded, "explicit, intentional processing [of Arabic numerals] is associated with individual differences in children's mathematical competence scores" (Bugden & Ansari, 2011, p. 40). In particular, children who could quickly determine which of two numerals (e.g., 2 4) was larger were more skilled than their slower peers at solving symbolic arithmetic problems. Geary and colleagues' prospective studies of mathematics achievement and risk of MLD confirm the importance of the fluent (i.e., fast and accurate) processing of the magnitudes associated with Arabic numerals (Geary, 2011; Geary, Bailey, & Hoard, 2009; Geary, Hoard, Nugent, & Bailey, 2012).

Example items from their measure, the number sets test, are shown in Figure 1, where the child's task is to quickly circle the combinations of nonsymbolic (i.e., small sets of objects) and symbolic (i.e., Arabic numerals) magnitudes that can be combined to match a target numeral (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Geary et al., 2009). The test is more complicated than the simple comparison of digit pairs, because it requires the combining of nonsymbolic or symbolic magnitudes, but nevertheless is dependent on the fluent processing of magnitudes associated with Arabic numerals. The utility of the measure was first demonstrated by Geary et al. (2009) who showed that performance in 1st grade correctly identified 2 out of 3 children who were diagnosed with MLD several years later and 9 out of 10 children who received no such diagnosis. The sensitivity and specificity of number sets performance was in fact better than that of 1st grade mathematics achievement scores, and did not predict reading achievement, indicating strong discriminant validity.

Subsequent research confirmed the importance of the competencies assessed by the test (Cirino et al., 2015; Geary, Hoard, Nugent, & Bailey, 2013; Fuchs et al., 2010; Fuchs, Geary, Fuchs, Compton, & Hamlett, 2014; Manolitsis, Georgiou, & Tziraki, 2013; Namkung & Fuchs, 2012). For example, Cirino et al. (2015) examined the specific deficits of 2nd grade children with MLD or reading difficulty (RD), and those with comorbid difficulties or no learning difficulties. Children with MLD regardless of their reading status showed particular deficits in number sets performance. Geary et al. (2013) found that number sets performance in 1st grade was an important component of children's broader number-systems knowledge that six years later predicted their performance on numeracy measures related to employability and wages in young adults, controlling domain general abilities, standardized mathematics achievement, and demographic factors. Despite these relations and the literature focusing more broadly on the importance of young children's fluency in the processing of the magnitudes associated with Arabic numerals (Bugden & Ansari, 2011; De Smedt & Gilmore, 2011; De Smedt et al., 2013; Lyons et al., 2012; Rousselle & Noël, 2007), little is known about the preschool quantitative competencies that support children's eventual fluency in this processing, and thus this is our focus here.

Current study

As noted, we focused on the preschool predictors of kindergarteners' performance on the number sets test. The test is useful in this regard because it captures fluency of processing the magnitudes of Arabic numerals that is an important component of children's early mathematical development, and predicts later mathematics achievement (e.g., Geary et al., 2013). Moreover, the measure also includes sets of problems that only involve combining nonsymbolic quantities (i.e., sets of objects) and mapping these to a target Arabic numeral. The mapping itself is

an important part of learning about Arabic numerals (Rousselle & Noël, 2007), and in this context (involving collections of objects in a numerical context) should in theory engage the ANS (Feigenson et al., 2004; Gallistel & Gelman, 1992). A second set of problems included a mix of nonsymbolic quantities and Arabic numerals or only Arabic numerals. For these problems, children's cardinal knowledge should in theory be at least as important if not more important than their ANS acuity. Children's cardinal knowledge should also be more important for these items than for the item set that only included nonsymbolic quantities.

To isolate preschool quantitative abilities that contribute to overall number sets performance, and performance on the different item sets, we simultaneously controlled for domain-general, preliteracy, and demographic factors that are correlated with mathematics achievement. These factors include executive functions, intelligence, alphabet recognition, speed of number naming, and parental background (Anderson et al., 2012; Clark, Pritchard, & Woodward, 2010; Geary, 2011; LeFevre et al., 2010).

Method

Participants

The sample included 112 children (51 boys) across two cohorts from the Columbia, MO public schools who were enrolled in Title I, a federally funded program designed to facilitate learning in three-to-five year olds at risk of school failure (see Chu, vanMarle, & Geary, 2013). All children included here completed all tasks across all three years of testing, and were part of a larger sample ($n = 158$) that began the study. The 46 children who did not complete all testing sessions, either because they had moved or were inattentive during a large number of sessions, had lower intelligence ($M = 92$, $SD = 19$) than the children who completed all testing sessions ($M = 97$, $SD = 16$), $F(1, 143) = 2.537$, $p = .113$. Thus, the results might not be generalizable to the

entire group of at-risk children, but should be reflective of more typical performance, because the final sample was average with respect to their mean intelligence.

Demographic information for the final sample was obtained by parent survey. Of the parents who completed the survey ($n = 93$), 84% reported that their children were of non-Hispanic ethnicity, 13% were Hispanic/Latino, and the ethnicity of remaining children was not reported. The racial composition was 60% White, 22% Black, 5% Asian, 11% more than one race, and the race of the remaining children was not reported. Income levels were: \$0-\$25k (39%), \$25-\$50k (26%), \$50-\$75k (22%), \$75-\$100k (11%), \$100-\$150 (1%), \$150k or more (1%). Forty-two percent of the respondents reported receiving food stamps and 9% reported receiving housing assistance. Maternal and paternal education levels were correlated ($r = .66$) and were thus collapsed into parental education categories corresponding to the highest level attained within the parent dyad; no information (17%), those with high school diplomas or less (28%), and those with at least some college experience (55%).

Quantitative tasks

ANS

The Panamath program (Halberda et al., 2008), a commonly used measure of ANS acuity, was used. Children in the first cohort received 24 test trials in year 1, and based on results for this cohort, we added 6 relatively easy trials for the second cohort and for year two. Each trial was presented on a standard laptop computer and consisted of blue and yellow dots contained within rectangles of their respective color. Children were asked to identify which set “had more dots.” Each set contained 5 to 21 dots and were displayed simultaneously for 2533ms to prevent verbal counting. Average dot size was 36 pixels and varied up to 20% to hold constant the total area of the dots for one half of the trials. The ratio of blue:yellow dots was randomly determined

on each trial and ranged from 1.29 to 3.38; the ratio for the 6 added items varied between 3.5 and 4.0. Percent accuracy was used as the measure of ANS acuity because it provides a more reliable estimate of sensitivity to relative quantity for young children than does the Weber fraction obtained from this task (Inglis & Gilmore, 2014).

Nonverbal calculation

Across 4 familiarization and 12 test trials, children watched the experimenter place discs under a mat and were then asked to predict the end result of adding or subtracting discs (Levine et al., 1992). Responses could include replicating the hidden set (after the addition or subtraction) or by verbally stating the final number of hidden discs. The test trials were presented in random order: 3-1, 2+2, 4-2, 1+3, 4-1, 4+1, 3+2, 1+4, 5-2, 5-3, 2+4, and 6-4. The score was the percent of correct responses for all completed trials.

Numerical recognition and comparison

We first showed children individual cards with the Arabic numerals from 1 to 15 in random order, one-at-a-time. The score was the total number of correctly named numerals. Following this, the cards of correctly identified numerals were shuffled together. On each trial (6 total), two cards were placed in front of the child, who was asked to indicate “which is bigger?” The associated numerical comparison score was the percent correct across the 6 trials.

Give-a-number

As described earlier, this task is commonly used to assess young children’s understanding of cardinal value (Sarnecka & Carey, 2008; Wynn, 1990). Children were asked to place exactly 1, 2, 3, 4, 5, or 6 cookies from a larger pile onto a plate to “feed” a puppet. The children were then asked to count the set aloud to ensure that the number of cookies placed on the plate matched the number they intended to provide. The task began with the set size of 1 and advanced to larger

sets until incorrect; if the number of cookies was incorrect, set size was reduced by 1. The highest set correctly provided on at least 2 out of 3 attempts was used as the measure of highest cardinal value known by the child, indicating their number knower-level (e.g., Lee & Sarnecka, 2011; Sarnecka & Carey, 2008); children who are not successful at providing any correct quantities are considered “pre-knowers”.

Point-to-x

As a second measure of cardinal knowledge, children were asked to “point to the picture that has X objects” (Wynn, 1990). Two blocks of 6 trials with ratios ranging from 0.5 to 0.67 (collections ranged from 1 to 10 in total) were presented. On each trial, children saw two collections of objects on a laptop display, presented horizontally. Selection of the smaller or larger collection, as well as the side containing the correct response, was counter-balanced across the sample. The score was the product of each trial’s score (1 = correct, 0 = incorrect) and the ratio of that trial’s collections. These products were then summed, providing a single score for each child that is weighted for difficulty.

Counting knowledge

Counting concepts (e.g., one-to-one correspondence) (Gelman & Gallistel, 1978) and awareness of essential and unessential features of counting (Briars & Siegler, 1984) were assessed. Thirteen trials were administered in which the child watched a puppet touch and count a line of checker pieces that alternated in color. After each counting event, the child indicated whether the count was “OK” or “Not OK and wrong”. Four types of trials were administered. Correct trials (3 total) showed a sequential and correct count from left-to-right. Right-left trials (3 total) were sequential and correct, but with the count procedure moving from right-to-left. Pseudo-error trials (4 total) showed a correct count from left-to-right, but with first counting all

checkers of just one color, a return to the far left, then a correct counting of the other colored checkers. Error trials (3 total) involved a sequential count from left-to-right, but with the first checker being counted twice. The score was the overall percent of trials identified correctly as “OK” (pseudo-error trials only) and “Not OK and wrong” (error trials); we focused on these trial types because they are correlated with arithmetic skills (Geary, Bow-Thomas, & Yao, 1992).

Number sets

As shown in Figure 1, domino-like stimuli were arranged in lines across a page, with each side of a domino ($1/2$ ” squares) containing different discrete magnitude-related information. The test included three stimuli types: one with only objects (e.g., squares), one with only Arabic numerals, and the third with objects on one side and an Arabic numeral on the other (18 pt font). A target sum of 5 is presented at the top of each page and children are told to circle all the dominoes that sum to the target value. Each page contains the following item types: 18 items matching the target value, 12 larger than the target, 6 smaller than the target, and 6 containing 0 or an empty square. The first page includes object-only items and the second includes Arabic numeral and a mix of Arabic numeral and object items.

The experimenter begins by explaining two items matching a target sum of 4; then, uses the target sum of 3 for practice. The test is then administered. The child is told to move across each line of the page from left to right without skipping any and to “circle any groups that can be put together to make the target number, 5.” Instructions place equal emphasis on speed and accuracy. A time limit of 60 seconds was chosen to avoid ceiling effects and to assess fluent recognition and manipulation of quantities for the target number. Signal detection methods, i.e. d-prime (z scores for hits – z scores for false alarms) (MacMillan, 2002), were used to assess

task fluency (Geary et al., 2007; 2009). Cronbach's alphas range from .70 to .90 for hits, misses, correct rejections, and false alarms (Geary et al., 2007).

Cognitive and achievement tests

Intelligence

Children completed the Receptive Vocabulary, Block Design, and Information subscales of the *Wechsler Preschool and Primary Scale of Intelligence-III* (WPPSI-III; Wechsler, 2002). Following standard procedures, scores were scaled and prorated to generate an estimate of full scale IQ; the reliability of the composite score range from .89 to .95 depending on age.

Executive functions

We administered the Conflict EF scale for children 2 to 6 years of age (Beck, Schaefer, Pang, & Carlson, 2011; Carlson, 2012). The task requires children to sort cards with different attributes into boxes according to certain rules (e.g., "big kitty" into "big kitty" box, "little kitty" into "little kitty" box). The rule was then reversed to create a conflict trial (e.g., "big kitty" into "little kitty" box). Subsequent trials required sorting by color or shape, and finally trials in which the sorting rule depended on the presence or absence of a border on the card. There were 7 levels for a total of 70 trials. Levels 1-4 each consist of 2 subsections, and levels 5-7 each consist of 10 trials. Children had to correctly answer 4 out of 5 trials in the subsections of levels 1-4 and 4 shape and 4 color (or 4 border and 4 non-border) trials in subsequent levels to advance onto the next subsection/level in the task. Testing began at level 2 for each child and continued until a level was failed. The number of correctly answered conflict trials was used as the child's score. The intraclass correlations for test-retest reliability range from .75 to .80 (Beck et al., 2011).

Speed of number naming

A rapid automatized naming task was used to assess processing speed (Denckla & Rudel, 1976; Mazzocco & Myers, 2003). The child is presented with 5 numbers to first determine if the child can read the stimuli correctly. After these practice items, the child is presented with a 5 X 10 matrix of instances of these same numbers, and is asked to name them as quickly as possible without making any mistakes. RT is measured via a stopwatch.

Preliteracy

The Upper-Case Alphabet Recognition subtest of the *Phonological Awareness Literacy Screening—PreK* (PALS: Invernizzi, Sullivan, Meier, & Swank, 2004) is a reliable indicator of later reading ability (Blatchford, Burke, Farquhar, Plewis, & Tizard, 1987) and was used here. The score was the total number of correctly identified letters.

Mathematics achievement

End of year mathematics achievement was measured with the *Test of Early Mathematical Ability-3* (TEMA-3: Ginsburg & Baroody, 2003). This test is nationally normed ($M = 100$, $SD = 15$); it includes items such as representing quantity through finger displays, counting, numeral comparison, and basic informal arithmetic. Testing began on the first item and ended once a child missed 5 consecutive items. Internal consistency is greater than .92, and alternate form and test-retest reliabilities are greater than .80. The mean performance of children included here was 92 ($SD = 15$) for the first year of preschool and 95 ($SD = 14$) for the second year.

Procedure

Children completed six 35-min testing sessions during each of the two years of preschool (see vanMarle, Chu, Li, & Geary, 2014): four quantitative sessions (two batteries, each administered once in the fall and once in the spring), one session for the cognitive and preliteracy tests between the fall and spring quantitative assessments, and an end-of-year TEMA-3

assessment. The IQ and preliteracy measures were assessed in the first year and executive functions and TEMA-3 in both years. The number sets test and the RAN were administered in the fall of kindergarten as part of a larger battery of number and arithmetic tasks. We used quantitative scores from the beginning (mean age = 3 years, 10 months; $SD = 3.9$ months; range 3 years, 3 months to 4 years, 3 months) and end of preschool (5 years, 2 months; $SD = 3.3$ months). The mean age at the time of the kindergarten assessment was 5 years and 9 months ($SD = 3.3$ months).

Analysis

We used both Bayes factors and standard regression techniques to make decisions about the best quantitative-task predictors of number sets fluency, that is, overall number sets scores, scores on object only items, and scores on mixed numeral and object items (Dienes, 2014; Gallistel, 2009; Raftery, 1995; Rouder & Morey, 2011; 2012; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007). The Bayes factors were computed for regression models (Liang, Paulo, Molina, Clyde, & Berger, 2008; Rouder & Morey, 2012) using the BayesFactor package (version 0.9.11-1) for R (Morey & Rouder, 2014). Alternative models are presented for each time point, with the first year of preschool represented as $MY1_m$, where m = the specific set of predictors in the model, and comparisons as $BY1_{mn}$, with B representing Bayes factor and m and n representing the compared models. $BY1_{m0}$ represents a contrast of the null model, with no predictors in the analysis, and alternative models. These analyses assess the likelihood of the data under different alternative models. The approach allows us to observe the degree of evidence in favor of alternative models, as contrasted with observing evidence in relation to the null model (Rouder & Morey, 2012; see Wagenmakers, Verhagen, Ly, Matzke, Steingroever, Rouder

&Morey, in press for discussion of the logical benefits of Bayesian inference over conventional null hypothesis significance testing).

As an example, if model MY_{1_1} consists of the predictors give-a-number and nonverbal calculation, and MY_{1_2} only consists of nonverbal calculation, the comparison of the models ($BY_{1_{12}}$; the subscripts represent the numerator and denominator, respectively, for calculating the odds ratio) is an odds ratio that represents how probable the data of MY_{1_2} are in explaining number sets performance as compared to MY_{1_1} . The same notations are used for the models including the year 2 preschool predictors, but with Y2 substituted in the model names. Generally, ratios that differ by a factor of 3 are considered suggestive evidence and those that differ by 10 are considered strong evidence (Raftery, 1995).

Bayes factors are more robust than standard linear regression in the selection of models that contain collinearity between predictors, as with our data set. The Bayes factor is higher when one of two highly correlated variables are included in relation to models containing both or none, providing the ability to compare the relative contribution of the predictors to the outcome measure (as odds ratios). In selecting models, we first compared all possible combinations of quantitative predictors. The approach is susceptible to type-I errors, and thus we also tested alternative models in which each predictor in the best model was dropped, one at a time, and the resulting model evaluated in terms of odds ratios. And, we identified the quantitative predictors that emerged in all or most of the top 30 Bayes models. These are variables that predicted number sets performance regardless of which other variables were included in the model. The best set of predictors along with the covariates were then regressed on number sets scores. For the Bayes analyses, give-a-number was included as a continuous variable, because of difficulties in modeling knower level within the BayesFactor package. For the regression analyses, five

knower-level categories were created (e.g., Le Corre & Carey, 2007; Lee & Sarnecka, 2011); due to small *ns*, pre-knowers ($n = 2$) were included with one-knowers, and five-knowers ($n = 3$) were included with four-knowers. In all, there were 19 pre- and one-knowers (hereafter one-knowers), 29 two-knowers, 19 three-knowers, 17 four- and five-knowers (hereafter four-knowers), and 28 CP-knowers.

Results

Beginning of preschool prediction of kindergarten number sets

Descriptive statistics for the predictors are shown in Table 1, and correlations among these predictors and kindergarten number sets performance in Table 2. The latter reveals that all predictors except the counting knowledge variables were correlated with one another and with overall number sets scores ($|r|s > .18$), a situation in which Bayes models provide an advantage over standard regression analyses.

The best Bayes models for the beginning of preschool predictors are shown in the top section of Table 3. The values of $BY1_{m0}$ are rather large, suggesting strong evidence in support of the models relative to the null. The best fitting model ($MY1_1$) included give-a-number, nonverbal calculation, and point-to-x, which also emerged in 30, 30, and 15 of the top 30 models, respectively. By comparison, numeral recognition (14/30), ANS acuity (12/30), numeral comparison (10/30), counting error (11/30) and pseudo-error trials (4/30) were less frequently selected. We compared the top model to more parsimonious models that excluded one variable at a time ("Excluded" column). Using Raftery's (1995) rule of thumb, models that are less than 33% as probable as the best fitting model provide evidence for the importance of the dropped predictor. For example, excluding point-to-x results in a model ($BY1_{21}$) that is 64.2% as probable as the best fitting model, and thus only provides weak evidence for the importance of this

predictor. As shown in Table 3, there is substantial evidence for the importance of nonverbal calculation (3.4% as probable without this predictor) and give-a-number (0.4%).

The top of Table 4 presents the regression analyses for these variables and the covariates. The results confirm the importance of number knower-level status, nonverbal calculation, and point-to-x; overall model, $F(12, 95) = 7.10, p < .001, R^2 = .47$. More specifically, we found that those children who had not yet achieved at least a three-knower status performed significantly worse on the number sets test in kindergarten, compared to CP-knowers; the effect for one-knowers was a trend ($p = .101$) and was significant for two-knowers ($p = .009$). Performance on point-to-x and nonverbal calculation also emerged as significant predictors, above and beyond the influence of knower level. Note that number knower-level was no longer significant when year 1 mathematics achievement scores were included as an additional control; point-to-x retained marginal significance ($\beta = 0.19, p = 0.058$), and nonverbal calculation remained significant ($\beta = 0.31, p = .004$).

We then redid the analyses for each of the two pages of the number sets test. The top Bayes model for the prediction of nonsymbolic processing to Arabic numeral mapping (object only items, page 1) included give-a-number and nonverbal calculation. Both predictors emerged in each of the 30 top Bayes models. With control of the covariates, nonverbal calculation ($\beta = 0.36, p = .009$) and knower-level status from four-knowers ($\beta = -0.91, p = .039$) and two-knowers ($\beta = -0.92, p = .018$) remained significant predictors of poor number sets performance. The performance of one-knowers was also below that of CP-knowers ($\beta = -0.92$) but the effect was not significant ($p = .142$). Only nonverbal calculation remained significant after the additional control of year 1 mathematics achievement ($\beta = 0.36, p = .007$). In the prediction of the symbolic and mix of symbolic and nonsymbolic items (number sets, page 2), the top Bayes

model included give-a-number, nonverbal calculation, and point-to-x, which emerged in 30, 26, and 24 of the top 30 models, respectively. With control of the covariates, two-knowers ($\beta = -0.53, p = .045$) performed worse than CP-knowers; one-knowers also performed worse than CP-knowers, but the effect was not significant ($\beta = -0.53, p = .153$). In this model, point-to-x ($\beta = 0.22, p = .017$) and nonverbal calculation ($\beta = 0.20, p = .037$) were also significant predictors. With additional control of mathematics achievement, point-to-x ($\beta = 0.21, p = .023$) and nonverbal calculation ($\beta = .21, p = .028$) remained significant, while knower-level did not ($ps > .30$).

End of preschool prediction of kindergarten number sets

As shown in the bottom section of Table 3, the best fitting model included nonverbal calculation, point-to-x, and numeral recognition (MY2₁), which were represented in 21, 30, and 21 of the top 30 models, respectively. All other variables were included in fewer than 12 models each. Dropping each variable from the top model provided substantial evidence for the importance of point-to-x (.3% as probable without this predictor), and only weak evidence for nonverbal calculation and numeral recognition. The linear regression results (bottom section of Table 4) were consistent with these findings, that is, point-to-x was highly significant ($\beta = .37, p = .004$) and there was a trend for nonverbal calculation ($\beta = .19, p = .069$); overall, $F(9, 102) = 8.40, p < .001, R^2 = .43$. Control of mathematics achievement eliminated the trend for nonverbal calculation ($\beta = .15, p = .18$), but point-to-x remained significant ($\beta = .36, p = .003$).

For the nonsymbolic items, the top Bayes model only included point-to-x; this variable emerged in all top 30 Bayes models. Point-to-x remained significant with control of the covariates ($\beta = .52, p < .0007$), including mathematics achievement ($\beta = .48, p < .002$). For the symbolic and mix of symbolic and nonsymbolic items, the best Bayes model included point-to-x,

numeral recognition, and nonverbal calculation. The Bayes model without point-to-x was only 2.6% as likely as the model with it, whereas the corresponding values for numeral recognition and nonverbal calculation were 41.2% and 90.1%, respectively. With control of the covariates, point-to-x remained significant ($\beta = .28, p < .012$), and there was a trend for nonverbal calculation ($\beta = .17, p = .088$), but numeral recognition was not significant ($\beta = -.02, p = .838$). Point-to-x remained significant with control of mathematics achievement ($\beta = .11, p = .013$), but nonverbal calculation did not ($p = .184$).

Discussion

There is now consistent evidence that children's symbolic number-system knowledge at the beginning of formal schooling predicts concurrent and later mathematics achievement (e.g., Cowan et al., 2011; Geary et al., 2013). The number sets test does not capture the full breadth of this knowledge, but it does require a core component of it; that is, the fluent processing and manipulation of the quantities associated with Arabic numerals (Bugden & Ansari, 2011). The test also provides separate measures of the fluency of mapping the quantities associated with collections of items to a target Arabic numeral, and a mix of these nonsymbolic items and Arabic numeral mappings to the target. The different item types should, in theory at least, differentially draw on the acuity of the ANS and symbolic cardinal knowledge, but this is not what we found.

Across analytic techniques (i.e., Bayes and regression) and the item types, performance on the give-a-number and nonverbal calculation tasks at the beginning of preschool always predicted number sets performance in kindergarten, even with control of executive functions, intelligence, alphabet recognition, speed of number naming, and parental background. At the beginning of preschool, there was some evidence that performance on the point-to-x task was an

important predictor of overall number sets performance and on performance on the mixed items but not the items that involved nonsymbolic to Arabic numeral mappings. At the end of preschool, and again regardless of analytic technique or item type, performance on the point-to-x task always predicted number sets performance in kindergarten, even with control of covariates. Performance on the nonverbal calculation and numeral recognition tasks emerged as predictors for overall number sets performance and on performance on the mixed items but not the items that only involved nonsymbolic to Arabic numeral mappings. Even for the former items, the overall importance on the nonverbal calculation and numeral recognition tasks was weaker than that of point-to-x. We discuss the implications of our results below, beginning with the ANS and then moving to cardinal knowledge, and finally to potential interactions among them.

ANS

As found in studies of mathematics achievement, children's ANS acuity at the beginning ($r = .32$, Table 2) and the end ($r = .39$) of preschool was correlated with their later performance on the number sets test (Fazio et al., 2014; Libertus et al., 2011; Mazzocco et al., 2011; Starr et al., 2013). Given these correlations and previous findings, we were somewhat surprised that our measure of ANS acuity did not emerge as a predictor of number sets performance, especially for the nonsymbolic set of items. One potential reason is that the ANS is sensitive to approximate quantity whereas the number sets test requires the processing of exact quantities, including the exact quantities associated with the object collections. Although the processing of item collections should trigger ANS activity (Feigenson et al., 2004), individual differences in this acuity, either at the beginning of preschool or end of preschool, did not emerge as critical, once children's symbolic knowledge was taken into consideration.

It may be that exact, symbolic number knowledge and nonsymbolic intuitions of number are unrelated (Matejko & Ansari, 2016; Negen & Sarnecka, 2014). It is also possible that preschoolers' ANS acuity indirectly contributes to later fluency of processing Arabic numerals, because preschoolers' learning of the cardinal principle appears to be supported, at least in part, by ANS acuity (Chu et al., 2015; vanMarle et al., 2014). In this view, the processing of Arabic numerals may still engage the ANS, in keeping with brain imaging studies (Ansari, 2016), but it is the explicit comparison and manipulation of the exact magnitudes associated with Arabic numerals that is critical to later processing fluency, in keeping with Bugden and Ansari (2011). So, the disconnection between symbolic number knowledge and ANS acuity may not occur at the neural level, but it does seem to occur at the functional level, that is, in the prediction of later mathematics achievement or associated symbolic competencies (e.g., in Arabic numeral comparison tasks).

It is also possible that the ANS influenced number sets performance through the nonverbal calculation task (Gelman, 2006; McCrink & Spelke, 2016). Indeed, end of preschool performance on this task was more strongly associated with ANS acuity than any other measure administered in this study, consistent with the proposal that the ANS embodies an intuitive understanding of aspects of arithmetic (Gallistel & Gelman, 1992). In this view, this intuitive knowledge might contribute to the nonsymbolic and symbolic addition required by the number sets test. However, the correlation between the ANS measure and nonverbal calculation was weaker at the beginning of preschool, when ANS acuity might be particularly helpful; that is, before children have much exposure to symbolic arithmetic. Also, children can solve these problems nonverbally, presumably based on the ANS, but also through counting. We do not have trial-by-trial information on how each item was solved, and so resolution of this issue will have

to await future studies, perhaps including tasks that require the integration of number-knowledge, counting, and implicit arithmetical knowledge (Zur & Gelman, 2004), as we elaborate below. However children solved these problems, competence in solving simple addition and subtraction problems at the beginning of preschool appears to contribute to their ability to fluently add quantities associated with sets of objects (as in the nonverbal calculation task), Arabic numerals, or a combination of them two years later.

Cardinal knowledge

Our results provide strong evidence that preschoolers who understand the cardinal value of number words (give-a-number) and the cardinal value associated with collections of items (point-to-x) are more fluent on the number sets test in kindergarten than their peers with less stable cardinal knowledge, independent of other quantitative skills, domain-general abilities, preliteracy knowledge, speed of number naming, and parental education. One might then argue that the number sets test is simply another measure of cardinality, but this cannot be the full story. Unlike the give-a-number and point-to-x tasks (Wynn, 1990), the number sets test and other measures of Arabic numeral processing that are predictive of concurrent or later mathematics achievement require the comparison or manipulation (e.g., addition) of two or more cardinal values (Bugden & Ansari, 2011; De Smedt & Gilmore, 2011; De Smedt et al., 2013). Cardinal knowledge of individual Arabic numerals would then be necessary but not sufficient for skilled performance on these measures. This follows logically, if we assume that the comparison and manipulation of two or more cardinal values is preceded by first learning the values of each of them separately.

The contribution here is the finding that individual differences in preschoolers' understanding of the cardinal value of Arabic numerals predicts fluent processing of two or more

cardinal values up to two years later. It is not that children need to be cardinal knowers two years before entering kindergarten, but they need to have taken several steps toward this end. Children who were one- and two-knowers at the beginning of preschool performed at least $0.5SDs$ lower on the number sets test, depending on item type, in kindergarten than did early cardinal principle knowers, controlling multiple other factors; we suspect the marginal findings for one-knowers was due to the smaller number of these children ($n = 19$) relative to two-knowers ($n = 29$), given the effect sizes across item types were very similar for these two groups. Based on previous studies, these children are at high risk for long-term difficulties learning mathematics in school (Geary et al., 2012). As noted, these results reinforce the importance of cardinal knowledge as a core part of the foundation for children's emerging mathematical competence (Chu et al., 2015; Negen & Sarnecka, 2014; vanMarle et al., 2014), and indicate that interventions to improve this knowledge in at risk preschool children will better prepare them for formal schooling.

Finally, the reader might wonder whether our conclusions are undermined by the finding that beginning of preschool performance on the give-a-number task was no longer a significant predictor of number sets scores when we controlled for year 1 mathematics achievement, and by the finding that give-a-number did not emerge as a critical predictor at the end of preschool. First, beginning of preschool give-a-number performance is highly correlated with mathematics achievement ($r = .65$, Table 2), and thus there is considerable collinearity. This level of correlation would in fact be expected if cardinal knowledge is central to children's mathematical development (vanMarle et al., 2014). Moreover, the mathematics achievement test includes many different types of items, including several that assess children's understanding of cardinality, and thus is not as informative as the give-a-number task. In other words, showing that earlier mathematics achievement predicts later number sets scores does not tell us much

about the foundation of number sets performance or performance on other measures of fluency of processing Arabic numerals, but findings for the give-a-number task do. Second, performance on the give-a-number task improved significantly during preschool and there were very few one- and two-knowers ($n = 7$) at the end of preschool, and thus the point-to-x task is likely more sensitive to individual differences in cardinal knowledge than give-a-number at this age.

Dynamic development of symbolic quantitative knowledge

Although our results confirm the importance of young children's learning of the cardinal values of number words and eventually Arabic numerals (Geary, 1994; Gelman & Gallistel, 1978; Wynn, 1990, 1992), they do not capture the dynamics of this learning. Sarnecka and S. Gelman (2006), for instance, demonstrated that young children appear to at least implicitly recognize that different number words beyond their knower level represent different quantities. So, a three knower does not explicitly know the quantities represented by 'five' or 'six' but they nevertheless recognize that the quantity signified by 'five' differs from the quantity signified by 'six', and both differ from more general signifiers of quantity, such as 'a lot'. Our tasks do not capture this transitional understanding of number words or the types of activities that move children from this transitional knowledge to CP knowers.

R. Gelman and colleagues have argued that children's knowledge of counting and number is guided by and fleshed out by their implicit understanding of arithmetic and their use of number and counting to achieve arithmetical goals (Gelman, 2006; Zur & Gelman, 2004). In other words, children's explicit understanding of number and other aspects of symbolic mathematics emerge from goal-directed, functional activities (Siegler & Jenkins, 1989), potentially constrained by knowledge implicit in the ANS (Gelman, 2006). Our tasks do not capture these dynamics, but the importance of the nonverbal calculation task in predicting later

number sets performance above and beyond the influence of children's cardinal knowledge is at least broadly consistent with this view. If correct, tasks that require the integration of number-knowledge and counting in the context of arithmetical problem solving (see Zur & Gelman, 2004) may be even more predictive of later fluency with numerals and mathematics achievement more generally than the tasks used in our study.

Limitations

Although our design was longitudinal and included controls that helped to eliminate alternative explanations for our findings, such as executive functions (Clark et al., 2010), the results are nevertheless correlational and thus intervention or experimental studies will be needed to fully evaluate our conclusions. Moreover, the sample was relatively small, especially for the regression analyses that divided it into five cardinal-knower levels, and thus our conclusions regarding the heightened risk of one- and two-knowers are in need of replication. The results are also in need of replication because the overall approach was exploratory. The preschool quantitative abilities assessed here were chosen based on previous research and thus not arbitrary; in fact, we expected ANS acuity and cardinal knowledge to emerge as predictors but were agnostic about the other predictors. Despite these limitations, our findings are consistent with the importance of children's early acquisition of the cardinal principle (Gelman & Gallistel, 1978; Le Corre & Carey, 2007) and achieving automaticity in accessing and manipulating cardinal knowledge in numerical contexts (Bugden & Ansari, 2011; Rousselle & Noël, 2007), and they also have clear implications for the design of early interventions.

References

- Anders, Y., Rossbach, H. G., Weinert, S., Ebert, S., Kuger, S., Lehl, S., & von Maurice, J. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Childhood Research Quarterly*, 27, 231-244.
- Ansari, D. (2016). Number symbols in the brain. In D. B. Berch, D. C., Geary, & K. Mann Koepke, K. (Eds.) (2016). *Development of mathematical cognition: Neural substrates and genetic influences* (pp. 27-50). San Diego, CA: Elsevier Academic Press.
- Beck, D.M., Schaefer, C., Pang, K., & Carlson, S.M. (2011). Executive function in preschool children: test—retest reliability. *Journal of Cognition and Development*, 12, 169-193.
- Blatchford, P., Burke, J., Farquhar, C., Plewis, I., & Tizard, B. (1987). Associations between preschool reading related skills and later reading achievement. *British Educational Research Journal*, 13, 15-23.
- Bonny, J.W., & Lourenco, S.F. (2013). The approximate number system and its relation to early math achievement: evidence from the preschool years. *Journal of Experimental Child Psychology*, 114, 375–388
- Brankaer, C., Ghesquière, P., & De Smedt, B. (2015). The effect of a numerical domino game on numerical magnitude processing in children with mild intellectual disabilities. *Mind, Brain, and Education*, 9, 29-39.
- Briars, D. & Siegler, R.S. (1984). A featural analysis of preschoolers’ counting knowledge. *Developmental Psychology*, 20, 607-618.
- Bugden, S. & Ansari, D. (2011). Individual differences in children’s mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition*, 118, 32-44.
- Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

- Carlson, S.M. (2012). *Executive Function Scale for Preschoolers. Test Manual*. Institute of Child Development, University of Minnesota.
- Chu, F. W., vanMarle, K., & Geary, D. C. (2013). Quantitative deficits of preschool children at risk for mathematical learning disability. *Frontiers in Developmental Psychology, 4*, 195.
- Chu, F.W., vanMarle, K., & Geary, D.C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology, 132*, 205-212.
- Cirino, P.T., Fuchs, L.S., Elias, J.T., Powell, S.R., & Schumacher, R.F. (2015). Cognitive and mathematical profiles for different forms of learning difficulties. *Journal of Learning Disabilities, 48*, 156-175.
- Clark, C. A., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental psychology, 46*, 1176-1191.
- Cowan, R., Cowan, Donlan, C., Shepherd, D., Cole-Fletcher, R., Saxton, M., Hurry, J. et al. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology, 103*, 786-803.
- Denckla, M. B., & Rudel, R. (1976). Rapid automatized naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14*, 471-479.
- De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*, 278-292.
- 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, 48-55.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2014.00781
- Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43, 1428-1446.
- 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, 53–72.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8, 307-314.
- Fuchs, L.S., Geary, D.C., Compton, D.L., Fuchs, D., Hamlett, C.L., Seethaler, P.M., Bryant, J.D., Schatschneider, C., (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology*, 46, 1731-1746.
- Fuchs, L.S., Geary, D.C., Fuchs, D., Compton, D.L., & Hamlett, C.L. (2014). Sources of individual differences in emerging competence with numeration understanding versus multidigit calculation skill. *Journal of Educational Psychology*, 106, 482-498.
- Gallistel, C.R. (2009). The importance of proving the null. *Psychological Review*, 116, 439-453.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation.

Cognition, 44, 43-74.

- Geary, D. C. (1994). *Children's mathematical development: Research and practical applications*. Washington, DC: American Psychological Association.
- Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, 37, 4-15.
- Geary, D.C. (2011). Cognitive predictors of individual differences in achievement growth mathematics: A five year longitudinal study. *Developmental Psychology*, 47, 1539-1552.
- Geary, D.C., Bailey, D.H., & Hoard, M.K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The Number Sets test. *Journal of Psychoeducational Assessment*, 27, 265-279.
- Geary, D. C., Berch, D. B., & Mann Koepke, K. (Eds.) (2015). *Evolutionary origins and early development of number processing* (Vol. 1, *Mathematical cognition and learning*). San Diego, CA: Elsevier Academic Press
- Geary, D. C., Bow-Thomas, C. C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology*, 54, 372-391.
- Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, 78, 1343-1359.
- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five year prospective study. *Journal of Educational Psychology*, 104, 206–223.

- Geary, D.C., Hoard, M.K., Nugent, L., & Bailey, D.H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE* 8(1): e54651. doi:10.1371/journal.pone.0054651
- Gelman, R. (2006). Young natural-number arithmeticians. *Current Directions in Psychological Science*, 15, 193-197.
- Gelman, R., & Gallistel, C.R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition*, 13, 343-359.
- Gersten, R., Clarke, B., Jordan, N.C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Council for Exceptional Children*, 78, 423-445.
- Ginsburg, H.P. & Baroody, A.J. (2003). *Test of early mathematical ability* (3rd edition). Austin, TX: Pro-ed.
- Gross-Tsur, V., Manor, O., & Shalev, R.S. (1996). Developmental dyscalculia: Prevalence and demographic features. *Developmental Medicine and Child Neurology*, 38, 25-33.
- Halberda, J., Mazocco, M., & Feigenson, L. (2008). Individual differences in nonverbal number acuity predict maths achievement. *Nature*, 455, 665-668.
- Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, 145, 147-155.
- Invernizzi, M., Sullivan, A., Meier, J., & Swank, L. (2004). *PreK teacher's manual: Phonological awareness literacy screening*. Charlottesville, VA: University of Virginia.

- Iuculano, T., Tang, J., Hall, C. W. B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science*, 11, 669-680.
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences USA*, 103, 10382-10385.
- Jordan, N.C., Kaplan, D., Ramineni, C., & Locuniak, M.N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, 45, 850-867.
- Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105, 395-438.
- Lee, M.D. & Sarnecka, B.W. (2011). Number-knower levels in young children: Insights from Bayesian Modeling. *Cognition*, 120, 391-402.
- LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, 81, 1753-1767.
- Levine, S.C., Huttenlocher, J., & Jordan, N. (1992). Development of calculation abilities in young children. *Journal of Experimental Child Psychology*, 53, 72–93.
- Liang, F., Paulo, R., Molina, G., Clyde, M.A. & Berger, J.O. (2008). Mixtures of g-priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103, 410-423.
- Libertus, M.E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with math abilities. *Developmental Science*, 14, 1292-1300.

- Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences USA*, 109, 18737–18742.
- Lyons, I.M., Ansari, D., & Beilock, S.L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General*, 14, 635-641.
- MacMillan, N.A. (2002). Signal detection theory. In: Wixted, J., Pashler, H., editors. *Stevens' handbook of experimental psychology, Vol. 4: Methodology in experimental psychology* (3rd Ed, pp. 43-90). New York: John Wiley & Sons.
- Manolitsis, G., Georgiou, G.K., & Tziraki, N. (2013). Examining the effects of home literacy and numeracy environment on early reading and math acquisition. *Early Childhood Research Quarterly*, 28, 692-703.
- Matejko, A.A. & Ansari, D. (2016). Trajectories of symbolic and nonsymbolic magnitude processing in the first year of formal schooling. *PLoS ONE* 11(3): e0149863. Doi: 10.1371/journal.pone.0149863.
- Mazzocco, M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, 6, e23749.
- Mazzocco, M. M. M., & Myers, G. (2003) Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia*, 53, 218-253.
- McCrink, K., & Spelke, E. S. (2016). Non-symbolic division in childhood. *Journal of Experimental Child Psychology*, 142, 66-82.
- Morey, R.D., & Rouder, J.N. (2014). BayesFactor 0.9.6, Comprehensive R Archive Network.

- Morgan, P.L., Farkas, G., & Wu, G. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, 42, 306-321.
- Namkung, J.M. & Fuchs, L.S. (2012). Early numerical competencies of students with different forms of mathematics difficulty. *Learning Disabilities Research and Practice*, 27, 2-11.
- Negen, J. & Sarnecka, B.W. (2014). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *British Journal of Developmental Psychology*, 33, 92-105.
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., . . . Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, 116, 33-41.
- Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163.
- Ritchie, S.J. & Bates, T.C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, 24, 1301-1308.
- Rouder, J.N. & Morey, R.D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, 18, 682-689.
- Rouder, J.N. & Morey, R.D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47, 877-903.
- Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, 16, 225-237.

- 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, 361-395.
- Rousselle, L. & Noël, M.-P. (2008). The development of automatic numerosity processing in preschoolers: Evidence for numerosity-perceptual interference. *Developmental Psychology*, 44, 544-560.
- Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition*, 92, 329-352.
- Sarnecka, B. W. & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, 108, 662-674.
- Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *The Quarterly Journal of Experimental Psychology*, 67, 271-280.
- Shalev, R.S., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine and Child Neurology*, null, 121-125
doi:10.1017/S0012162205000216
- Siegler, R. S., & Jenkins, E. (1989). How children discover new strategies. Hillsdale, NJ: Erlbaum.
- Siegler, R. S., & Shrager, J. (1984). Strategy choice in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229-293). Hillsdale, NJ: Erlbaum.
- Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts

- mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 18116-18120.
- Vanbinst, K., Ansari, D., Ghesquière, P., & De Smedt, B. (2016). Symbolic numerical magnitude processing is as important to arithmetic as phonological awareness is to reading. *PLoS ONE* *11*(3): e0151045. Doi: 10.1371/journal.pone.0151045.
- vanMarle, K., Chu, F.W., Li, Y., & Geary, D.C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Developmental Science*, *17*, 492-505.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problem of p values. *Psychonomic Bulletin and Review*, *14*, 779-804.
- Wagenmakers, E.-J., Verhagen, A.J., Ly, A., Matzke, D., Steingroever, H., Rouder, J.N., & Morey, R.D. (in press). The need for Bayesian hypothesis testing in psychological science. In Lilienfeld, S.O., & Waldman, I. (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions*. New York: John Wiley and Sons.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155-193.
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*, 220-251.
- Zur, O., & Gelman, R. (2004). Young children can add and subtract by predicting and checking. *Early Childhood Research Quarterly*, *19*, 121-137.

Table 1

Means and standard deviations for study predictors in the beginning and end of preschool.

Variable	Beginning of Preschool		End of Preschool	
	Mean	S.D.	Mean	S.D.
Mathematics Achievement	92.26	14.94	94.88	14.14
Executive Functions	31.57	13.88	45.08	13.29
Intelligence	97.30	16.23	--	--
Pre-Literacy	12.62	9.24	--	--
ANS	66.6%	16.05	84.2%	17.73
Give-a-Number	3.34	1.87	4.07	1.9
Nonverbal Calculation	23.01%	16.24	43.5%	21.75
Point-to-X	2.14	0.54	2.77	0.55
Numerical Recognition	5.68	4.21	9.42	4.05
Numerical Comparison	59.4%	22.22	82.3%	21.79
Counting Knowledge-Pseudo Error	76.8%	31.4	64.9%	35.97
Counting Knowledge-Error	24.3%	34.89	61.2%	36.27

Note. ANS = percent correct on the Panamath program (Halberda et al., 2008).

Table 2

Correlations among the beginning(top) and end (bottom) of preschool quantitative measures, covariates, and number sets fluency in kindergarten.

Variables	Beginning of Preschool Correlations												
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Mathematics achievement	1												
2. Executive functions	.42	1											
3. Intelligence	.53	.41	1										
4. Pre-literacy	.57	.27	.33	1									
5. ANS	.44	.37	.34	.16	1								
6. Give-a-number	.65	.45	.47	.43	.36	1							
7. Nonverbal calculation	.27	.28	.29	.17	.18	.29	1						
8. Point-to-x	.31	.28	.25	.21	.28	.30	.30	1					
9. Numeral recognition	.39	.21	.28	.48	.22	.26	.26	.26	1				
10. Numerical comparison	.13	.06	.06	-.02	.28	.16	.06	.27	.06	1			
11. Counting knowledge-pseudo error	.06	-.12	.08	.05	.05	-.07	.09	.06	.17	-.08	1		
12. Counting knowledge-error	.01	.15	-.10	-.04	-.05	.11	-.04	-.08	-.12	.03	-.52	1	
13. Number sets (d')	.56	.27	.37	.42	.32	.47	.43	.36	.32	.18	.01	-.08	1
End of Preschool Correlations													
1. Mathematics achievement	1												
2. Executive functions	.35	1											
3. Intelligence	.39	.48	1										
4. Pre-literacy	.59	.34	.33	1									
5. ANS	.42	.28	.34	.34	1								
6. Give-a-number	.45	.26	.31	.35	.48	1							
7. Nonverbal calculation	.42	.33	.21	.28	.47	.37	1						
8. Point-to-x	.43	.40	.36	.39	.49	.48	.41	1					
9. Numeral recognition	.69	.30	.34	.63	.39	.51	.32	.45	1				
10. Numerical comparison	.39	.15	.29	.29	.36	.21	.26	.36	.40	1			
11. Counting knowledge-pseudo error	.04	.01	-.07	.12	.07	-.01	.06	.02	.04	.13	1		
12. Counting knowledge-error	.15	.22	.11	.15	.28	.27	.25	.20	.29	.16	-.15	1	
13. Number sets (d')	.53	.41	.37	.42	.39	.29	.41	.54	.43	.35	-.04	.19	1

Note. $n = 112$. ANS = percent correct on the Panamath program (Halberda et al., 2008)

Table 3

Bayes factor analysis of beginning and end of preschool predictors of kindergarten number sets performance

Model Beginning of Preschool	BY1_{m0}	Excluded	BY1_{m1}
MY1 ₁ GiveN + NVC + PointX	1.43×10^7	---	1
MY1 ₂ GiveN + NVC	9.20×10^6	PointX	.642
MY1 ₃ GiveN + PointX	4.87×10^5	NVC	.034
MY1 ₄ NVC + PointX	5.83×10^4	GiveN	.004
Model End of Preschool	BY2_{m0}	Excluded	BY2_{m1}
MY2 ₁ NVC + PointX + NumRecognition	1.18×10^8	---	1
MY2 ₂ PointX + NumRecognition	6.34×10^7	NVC	.537
MY2 ₃ NVC + PointX	4.75×10^7	NumRecognition	.402
MY2 ₄ NVC + NumRecognition	3.01×10^6	PointX	.003

Note. GiveN = Give-a-number; NVC = Nonverbal calculation; PointX = Point-to-x;

NumRecognition = Numeral recognition. The variables in the Excluded column indicate the variable excluded from MY1₁ (top) and MY2₁ (bottom).

Table 4

Linear regression results for preschool predictors of kindergarten number sets performance.

Beginning of Preschool	Estimate	S.E.	t value	Pr(> t)
Intercept	0.29	0.33	0.87	0.386
Parent education				
Unknown	0.00	--	--	--
H.S. degree or less	0.04	0.30	0.13	0.895
At least some college	0.17	0.28	0.60	0.550
Intelligence	0.12	0.12	0.99	0.326
Executive functions	-0.01	0.11	-0.11	0.910
Pre-literacy	0.20	0.12	1.68	0.096
Processing Speed	-0.24	0.10	-2.42	0.018
Give-a-number				
CP-Knower Level	0.00	--	--	--
4 Knower Level	-0.54	0.34	-1.58	0.118
3 Knower Level	-0.22	0.34	-0.66	0.510
2 Knower Level	-0.80	0.30	-2.68	0.009
1 Knower Level	-0.69	0.42	-1.65	0.101
Point-to-x	0.21	0.10	2.07	0.041
Nonverbal calculation	0.30	0.11	2.82	0.006
End of Preschool	Estimate	S.E.	t value	Pr(> t)
Intercept	0.14	0.24	0.57	0.569
Parent education				
Unknown	0.00	--	--	--
H.S. degree or less	-0.23	0.30	-0.77	0.445
At least some college	-0.17	0.29	-0.60	0.551
Intelligence	0.15	0.11	1.31	0.192
Executive functions	0.13	0.11	1.15	0.252
Pre-literacy	0.15	0.12	1.19	0.236
Processing Speed	-0.18	0.10	-1.69	0.093
Point-to-x	0.37	0.12	2.97	0.004
Nonverbal calculation	0.19	0.11	1.84	0.069
Numeral recognition	0.06	0.13	0.45	0.651

Note. The regression analysis was conducted after excluding 4, pre- and one knower-level children who's scores on the number sets test were 1 S.D. away from the nearest score in that knower-level grouping. With the four children included, the β s of the regression were pre- and one-knowers: -0.46 ($p = .26$); two-knowers: -0.76 ($p = .02$); three-knowers: -0.28 ($p = .44$); four- & five-knowers: -0.46 ($p = .20$).

Figure 1: Example items from the *Number Sets Test*.