# How are approximate number, estimation, and arithmetic related in development? Evidence from a 3 year longitudinal curriculum intervention

Jessica Sullivan[1], Michael Frank[2], & David Barner[3]

[1]Dept. of Psychology, Skidmore College

[2]Dept. of Psychology, Stanford University

[3] Dept. of Psychology, University of California, San Diego

Word Count: 5,469

CONTACT:

Jessica Sullivan

Department of Psychology, Skidmore College

815 N. Broadway, Saratoga Springs, NY, 12866

t: (518)-580-5310; f: 518-580-5319

e-mail: jsulliv1@skidmore.edu

Abstract

What is the relationship between nonverbal, approximate number abilities and exact mathematics? A number of studies have found a correlation between the Approximate Number System (ANS) and early math success. Others have failed to replicate this finding, however, and few have addressed whether the relationship is causal. In the present study, we analyzed a large ($N = 204$) 3-year longitudinal dataset from a successful math intervention. We asked whether (a) changes to math performance induced via intervention caused changes to approximate number skill; and (b) whether correlations between math performance and approximate number skill were uniquely numerical, or instead were better explained by other differences in cognitive ability. While we replicated past work finding correlations between approximate number skill and math success, we found no evidence that changes to math performance induced changes to approximate number skill, and we found only limited evidence that approximate number skill predicted math performance above and beyond other cognitive skills like working memory or general intelligence.

Keywords: numerical cognition, math, mental abacus, cognitive development

Beginning early in infancy, humans can represent approximate numerical quantities nonverbally (Xu & Spelke, 2000), using what is sometimes called the "Approximate Number System" (ANS) or the "number sense." The ANS is used to represent and compare numerical magnitudes. It follows Weber's law, such that the ratio of any two numerical quantities determines the precision with which they can be differentiated (for review, see Dehaene, 1997). A number of recent studies report that individual differences in ANS acuity are related to mathematics achievement, such that individuals with greater numerical acuity also perform better on standardized math tests, the SAT, and a host of other math measures (Halberda, Mazzocco, & Feigenson, 2008; Halberda et al., 2012; Chen & Li, 2014). These results are exciting for at least two reasons. First, they suggest a link between the evolutionarily ancient ANS and the more recent human innovation of symbolic arithmetic, thus potentially providing insight into the origins of mathematical thought. Second, they suggest that tests of ANS acuity may be helpful in designing diagnostic and intervention tools for early math difficulties (Park & Brannon, 2013; 2014; Starr, Libertus, & Brannon, 2013), perhaps even before children begin formal math training.

In total, more than a dozen studies have reported some correlation between the ANS and symbolic math, and this correlation often survives the addition of non-numerical control predictors, like verbal SAT score, IQ, and spelling ability (e.g., Anobile, Stievano, & Burr, 2013; Bonny & Lourenco, 2013; Desoete et al., 2012; deWind & Brannon, 2012; Gilmore et al., 2010; Halberda, Mazzocco, & Feigenson, 2008; Halberda et al., 2012; Mazzocco et al., 2011a;b; Libertus et al., 2011; Libertus, Odic, & Halberda, 2012; Libertus, Feigenson, & Halberda, 2013; Piazza et al., 2010;

Starr et al., 2013). Further, Park and Brannon (2013; 2014) have provided evidence that training on non-verbal number tasks can lead to improvements to math performance in adults, raising the possibility that the ANS is foundational to mathematical learning, not merely an interesting correlate.

The empirical picture is more complicated, however: Many other studies have found the relation between ANS acuity and symbolic math ability to be negligibly small or even absent, especially when controlling for other non-numerical cognitive skills like inhibitory control, symbolic number knowledge, and non-numerical quantity comparison (e.g., Fuhs & McNeil, 2013; Gilmore et al., 2013; Göbel et al., 2014; Holloway & Ansari, 2009; Kolkman et al., 2013; Nosworthy et al., 2013; Price et al., 2012; Sasanguie, De Smedt, Defever, & Reynvoet, 2012; Sasanguie, Defever, Maertens, & Reynvoet, 2014; Sasanguie et al., 2013; Tibber et al., 2013; Wei et al., 2012). Further, at least one study that trained children's nonverbal number skill failed to demonstrate that training to the ANS improved math skill (Obersteiner, Reiss, & Ufer, 2013). These discrepant findings raise important questions about the nature and practical significance of any relationship between the ANS and symbolic math success (for review, see De Smedt, Noel, Gilmore, & Ansari, 2013).

One way to adjudicate between these discrepant findings is via meta-analysis. For example, one recent meta-analysis demonstrated that – across a wide range of study methodologies and samples (36 independent samples) – ANS acuity explained substantial variability in symbolic math achievement (Chen & Li, 2014). However, fully half of the studies included in the meta-analysis did not control for participants' non-numerical cognitive capacities, and the vast majority of those that did only controlled for participants' linguistic ability. This inconsistent selection of control tasks is a problem, because tasks typically used to measure ANS acuity – e.g., dot array comparison – likely draw on a host of non-numerical cognitive capacities, like working memory, non-numerical quantity representation, and inhibitory control. Critically, each of these cognitive capacities has also been shown to support early math achievement, making it possible that they directly mediate correlations between ANS acuity and math achievement (Alloway & Passolunghi, 2011; Clark, Pritchard, & Woodward, 2010; DeStefano & Lefevre, 2004; Geary, 2011; Gilmore et al., 2013; Hornung, Schlitz, Brenner, & Martin, 2014; Lourenco, Bonny, Fernandez, & Rao, 2012; Thompson, Nuerk, Moeller, & Cohen Kadosh, 2013). Thus, although reported correlations between ANS acuity and math achievement may be due to a unique relationship between verbal and nonverbal numerical abilities, it is also possible that other, non-numerical perceptual and cognitive capacities explain the reported correlations.[1]

In the present study, we tested whether the ANS bears a meaningful causal relationship to mathematics achievement, by assessing its longitudinal predictive power relative to a large battery of other cognitive measures. Specifically, we analyzed publicly available data from a three-year longitudinal randomized controlled trial of a math intervention in 204 2nd through 5th graders (*n* = 204). This RCT manipulated the mathematics ability of an experimental group via supplemental training in a popular mental arithmetic technique called "mental abacus" (Barner et al., in press). Mental abacus training involves teaching participants to perform arithmetic calculations using an

[1] Similar challenges confront the interpretation of studies that compare individuals with different levels of formal mathematics training (e.g., Castronovo & Göbel, 2012; Lindskog, Winman, & Juslin, 2014; Pica, Lemer, Izard, & Dehaene, 2004; Nys et al., 2013).

abacus, and, at more advanced levels, removing the physical abacus and asking users to visualize the abacus and calculations using this mental representation (Frank & Barner, 2011; Hatano, Miyake, & Binks, 1977; Stigler, Chalip, & Miller, 1986). In the RCT, abacus training improved children's arithmetic abilities above and beyond additional training in a standard math curriculum. Therefore, these data allowed us to test whether inducing changes to math skill caused measurable changes in ANS acuity. Also, the study included longitudinal measures of spatial working memory, verbal working memory, mental rotation ability, as well as a test of general intelligence (Raven's Progressive Matrices). These measures allowed us to assess the predictive relation between ANS acuity and math achievement while concurrently controlling for a large battery of non-numerical cognitive capacities.

Finally, we tested the relationship between numerical estimation performance (the ability to label arrays of dots with number words) and math achievement. Given that ANS representations are known to be linked to number words in the verbal count list, it is possible that the strength and precision of this link mediates the relation between ANS acuity and symbolic math performance (Libertus, Odic, Feigenson, & Halberda, 2015). Since estimation ability captures the translational process between symbolic and non-symbolic number representations (see Sullivan & Barner, 2014a, for discussion), we also tested whether children's estimation ability was uniquely related to math performance (Gunderson, Ramirez, Beilock, & Levine, 2012; Kolkman et al., 2013; Moore & Ashcraft, 2015; Siegler & Booth, 2004; Booth & Siegler, 2006; 2008).

To summarize, we used data from our longitudinal math intervention to test four main questions. First, we asked whether our math intervention caused changes to the

ANS. Second, we asked whether ANS ability concurrently predicted math performance and, if so, whether any correlation between ANS ability and math performance survived the addition of non-numerical control tasks. Third, we asked whether ANS ability predicted growth in math skill over time. Finally, we asked whether estimation ability also predicted math performance.

## Method

### *Participants*

Data were obtained from a previous study by Barner et al. (in press) at https://github.com/langcog/mentalabacus. Participants were 204 children from a charitable school in Gujarat, India. Children spoke English (the language of instruction at their school), and most children also spoke an additional language (Guajarati and Hindi were the most common). Children came from either Muslim (41%) or Hindi (59%) families. Family income was low, with 80% of children coming from families earning around $2000 USD per year. Children were between of 5 and 7 years of age at the time of enrollment (Year 0).

### *Math Measures*

Children received several measures of math competence, including the Woodcock-Johnson III Computation test, the Math Fluency subtest of the Weschler Individual Achievement Test (WIAT-III), and in-house tests of arithmetic and place value understanding. Also, children's math grades were available, as reported by their school. Detailed descriptions of these measures are available in the Supplementary Materials of Barner et al. (in press).

### *Intervention*

As reported in Barner et al. (in press), all children participated in standard math class throughout the 3 years of the study. In addition, at enrollment in Year 0, children were randomly assigned to classrooms where they received one of two supplemental curricula: A standard international math curriculum *or* mental abacus instruction. A total of 100 children were assigned to the mental abacus group and 104 to the control group. Of these 204 children, the majority provided data for every year of testing (Years 0-3; Control group $n$ = 88; Mental Abacus group $n$ = 99).

To test whether differences in math training are responsible for changes to the ANS and estimation ability, we compared the ANS and estimation abilities of children who completed the mental abacus intervention to the control group.

*ANS and Estimation Tasks*

Children's ANS acuity was assessed using a 10 minute computerized task. Two arrays of black dots were presented simultaneously on a gray background; the two arrays were separated by a vertical black line. Half of the trials controlled for total surface area across the arrays; the other half of trials controlled for item size (Dehaene et al., 2005). The correct answer was on the left 50% of the time.

Arrays were visible for 1000 ms and were followed by a 300 ms white noise mask image. Children were instructed to indicate which array was more numerous by pressing the Z (which was covered with a left arrow) or M (which was covered by a right arrow) key. The experiment was self-paced, and children pressed the space-bar to progress to the next trial. To ensure that children attended to each trial, two beeps were presented via headphones immediately prior to the presentation of the arrays.

Trials were presented in blocks of 8. Within each block, the ratio of items in the two sets remained constant; all children started with a 4:5 ratio. Within each block, the numerical magnitudes of the arrays varied substantially (e.g., 16 vs. 20; 80 vs. 100). In order to succeed on a given block, the child needed to get 6 out of 8 trials correct. Side of the correct response was pseudo-randomly ordered so that alternating responses or consistent choices of "left" or "right" would lead to failure of the block. If they succeeded on a block, they moved to the next hardest ratio (e.g., 5:6), while if they failed, they moved to the next easiest ratio (e.g., 3:4).

Children's estimation ability was tested by asking them to estimate the number of dots on a screen; arrays were randomly generated and contained black dots on a gray background. The number of dots ranged from 3-120, and dot size and total area of the array varied across trials. Children viewed each array for 400 ms, and then entered their numerical estimate on a keypad. Prior to beginning the task, children completed a keypad typing training session to ensure that all children could appropriately use the keypad. Total task duration was 10 minutes.

*Control Tasks*

Children were tested on a battery of control tasks, again described in detail at in the Supplemental Materials of Barner et al. (in press). This battery included two computerized measures of working memory: (1) a test of verbal working memory, which required recalling a sequence of syllables in the correct order; and (2) a test of spatial working memory, which required recalling a sequence of dot-locations in the correct order. Children also completed a paper-and-pencil task that measured mental rotation ability. Finally, children completed a subset of Raven's Progressive Matrices.

## Results

Before presenting our main analyses, we first describe how data were used to construct measures of ANS acuity, estimation ability, and mathematics achievement.

### Measures & Descriptive Statistics

**ANS and estimation measures**. Each child's ANS acuity was measure for each year (Year 0, 1, 2, and 3) as a Weber fraction (using the method described by Halberda, Mazzocco, & Feigenson, 2008), where greater ANS acuity is indicated by a smaller Weber fraction ($M_{acuity}$ Y0 = .37; Y1 = .18; Y2 = .15; Y3 = .14).

We also constructed several measures of estimation ability, using data from Years 1, 2, and 3 (Year 0 estimation data were not collected). First, we tested the internal consistency of children's estimates. To assess consistency we used two measures: ordinality and linear $r^2$, described below.

Ordinality captures the extent to which a child's estimates are ordered consistently, and can be measured by calculating the proportion of trials on which the child gave estimates in the correct direction relative to previous estimates. For example, if a smaller number of dots was shown on trial *n* than on trial *n-1*, a child's estimate was labeled as ordinal if their estimate is smaller on trial *n* than on trial *n-1*. The average rate of ordinal responding for each year was Y1 = .80; Y2 = .80; Y3 = .80. Previous work has shown that children can provide ordinal estimates long before they provide accurate estimates, suggesting that this measure might capture children's early structural knowledge of the relation between the verbal and nonverbal number systems (Sullivan & Barner, 2014a; 2014b). Recent work has even suggested that the understanding of ordinality mediates the link between the ANS and math achievement (Lyons & Beilock, 2011).

The linear $r^2$ measure of internal consistency represents the amount of variability in estimation performance that can be accounted for by knowing the number of dots a child was estimating. In other words, this value represents the extent to which the relation between a child's estimate and the number of dots that they saw can be described by a linear function (in previous work, this has been referred to as the "linearity" of children's estimates; e.g., Booth & Siegler, 2006). To calculate linear $r^2$, we constructed a linear regression predicting each child's estimates from the number of dots presented, and then reported the linear $r^2$ of the line (Y1 = .37; Y2 = .35; Y3 = .36). Importantly, a high linear $r^2$ score does not necessarily indicate that a child provided accurate estimates, but rather that the child's estimates were internally consistent (for example, one could imagine a child who overestimated small numbers, underestimated large numbers, and yet still provided estimates that were perfectly linear). Unlike ordinality, which only captures the internal consistency of the *ordering* of estimates, in order to have a high linear $r^2$ value, children must also be internally consistent in the *relative distance* between estimates. In number-line estimation tasks, linear $r^2$ has been shown repeatedly to correlate with symbolic math performance (Gunderson, Ramirez, Beilock, & Levine, 2012; Kolkman et al., 2013; Moore & Ashcraft, 2015; Siegler & Booth, 2004; Booth & Siegler, 2006; 2008).

In addition to these two measures of internal consistency of estimates we also calculated the accuracy of estimates via the Proportion Absolute Error (PAE) which represents the absolute value of the deviation of an estimate from the actual number presented, divided by the number presented ($M_{\text{PAE}}$: Y1 = .71; Y2 = .71; Y3 = 70). PAE has been shown previously to predict math performance on standardized tests (Siegler & Booth, 2004, Castronovo & Göbel, 2012; although not across all studies: see Booth & Siegler, 2006), addition/subtraction performance (Link, Nuerk, and Moeller, 2014; Moore & Ashcraft, 2015 [addition only]), and mental arithmetic (Lyons, Price, Vaessen, Blomert, & Ansari, 2014). Thus, if possessing highly accurate and stable mappings between number words and nonverbal representations of number is important to math success (e.g., in the case that children actually recruit the ANS to check or compute symbolic math calculations), then PAE should be related to math performance.

To provide an approximate measure of the reliability of each our estimation and ANS measures, we predicted each year's data from the previous year's data; we report these Pearson correlation coefficients and significance level in Table 1.

*Table 1.* Year-to-year reliability for each of our estimation and ANS measures. * indicates *p*<05, ** indicates *p*<.01, and *** indicates *p*<.001.

| | Ordinality | PAE | Linear *r*^2 | ANS |
|---|---|---|---|---|
| Year 0-1 | n/a | n/a | n/a | 0.232** |
| Year 1-2 | 0.213** | 0.278** | 0.191** | 0.191* |
| Year 2-3 | 0.333*** | 0.542*** | 0.379*** | 0.444*** |

**Math measures.** We created a composite of the standardized math test scores (WIAT and WJ-III) by averaging scores on the two to compute the proportion of correct responses across measures. This composite showed improved mathematics performance across each year of testing ($M_{\text{standardized}}$: Y0 = .20; Y1 = .31; Y2 = .43; Y3 = .54).

We also created a single composite measure that represented mathematical ability. To derive this measure, we conducted a Principal Components Analysis on all of our symbolic math measures (WIAT, WJ-III, arithmetic, place value, math grades). We then took the first principal component (PC1) as a measure of the primary shared variance between these tasks. We then predicted this measure – PC1 – from ANS and estimation performance. Our goal in creating these two composite measures was to reduce the dimensionality of our analyses and avoid the issue of attempting to analyze five different but highly correlated measures of symbolic math.

**Analyses**

**Effect of intervention on ANS and estimation.** We first compared the ANS acuity and estimation ability of children who received abacus instruction to that of children in the control group. As reported by Barner et al. (in press), abacus training had a significant impact on mathematics achievement, when measured by the WJ-III and the in-house arithmetic battery. Our analysis therefore allowed us to ask whether improving math performance caused improvements to ANS and estimation outcomes. To do this, we created a mixed effects linear regression model predicting ANS acuity from Year, Intervention Condition (abacus vs. no abacus), and their interaction. We also added participant-level random effects of Year, capturing individual children's growth over time. For this and all analyses, data and code are available at: https://github.com/langcog/jesstimation. If improving at math causes improvements to ANS acuity, then such improvements should result in a significant Year by Intervention interaction, such that children who received abacus training (and therefore got better at math) showed larger improvements in and acuity over time than children in the control group.

Our fitted model showed an effect of Year, suggesting that children's ANS acuity improved over time (*B* = -.07, *SE* = .03; the negative coefficient captures children's smaller Weber fractions). However, we found no effect of Intervention Condition (*B* = -.02, *SE* = .01) or interaction of Intervention Condition and Year (*B* = .006, *SE* = .007, *p* > .3), and planned *t*-tests also revealed no effects of abacus training on ANS in Years 1 (*t*(179) = -.20, *p* = .84), 2 (*t*(183)=-.34, *p* = .74), or 3 (*t*(183) = -1.26. *p* =.26; Fig. 1).

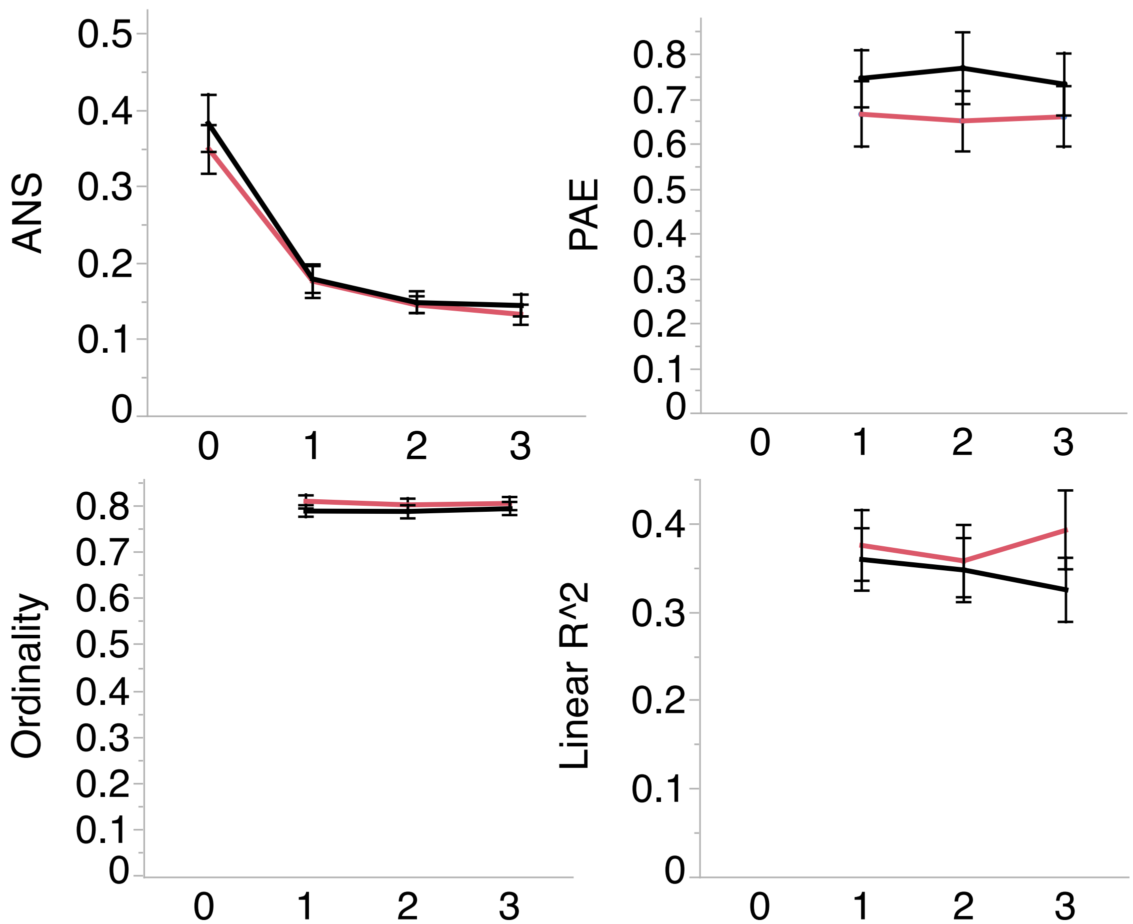


Figure 1. Difference across intervention conditions in estimation performance by Year (1, 2, and 3 – we did not test estimation in Year 0), and for ANS (Years 0, 1, 2, and 3). Red line indicates children who learned abacus; Black line indicates children who were in the control group. For our Linear $r^2$ and Ordinality measures, larger numbers indicate better performance; for Proportion Absolute Error and ANS measures, smaller numbers indicate better performance. Error bars are 95% Confidence Intervals.

We next tested whether our intervention influenced estimation performance. Because we did not have Year 0 (baseline) data for estimation, we could not assess with certainty whether abacus training caused changes to estimation. However, we were able to test whether there were differences in estimation performance between the abacus and control group during Years 1, 2, and 3.

No estimation measure showed consistent (e.g., across more than one year) differences between the abacus and control group. Also, when correcting for multiple comparisons, no *p*-value reached significance. For PAE, there was no effect of abacus training in Year 1 ($t(182) = -1.64$, $p = .10$), a significant, uncorrected, effect in Year 2 ($t(185) = -2.21$, $p = .03$; Bonferroni $p = .09$), and no effect in Year 3 ($t(184) = -1.47$, $p = .14$). For Ordinality, there was a significant, uncorrected, effect in Year 1 ($t(182) = 2.15$, $p = .03$; Bonferroni $p = .10$), and no effect in Years 2 or 3 (year 2: $t(185) = 1.39$, $p = .17$; year 3: $t(184) = 1.12$, $p = .26$). For Linear $r^2$, there were no effect in Years 1 or 2 (Year 1: $t(182) = .598$, $p = .55$; Year 2: $t(185) = .38$, $p = .71$), though there was a significant, uncorrected, effect in Year 3 ($t(184) = 2.34$, $p = .02$; Bonferroni $p = .06$). Although there were trends indicating a relation between Intervention Condition and estimation performance (e.g., see Fig. 1), none of these comparisons reached significance when correcting for the number of comparisons conducted. Further, there were no overall trends of improvement in estimation performance over time in relation to abacus training.

To summarize: These analyses suggest that an intervention that improved math performance did not significantly improve ANS acuity. Also, they failed to find evidence for a benefit of training to estimation performance, though we lacked pre-intervention estimation data that could allow us to definitively test whether the intervention affected estimation. These findings militate against one class of possible causal relation between the ANS and mathematics achievement – e.g., whereby changes in math achievement cause changes that result in greater ANS acuity. One limitation of these findings, however, is that mental abacus differs in many ways from other methods for improving math performance: MA training appears to recruit cognitive skills not directly related to math (Barner et al., in press; Frank & Barner, 2011), and MA calculations activate *both* regions of the brain associated with numerical processing *and* several additional regions (like those associated with visuospatial processing; Du et al., 2013). It therefore remains possible that because of the properties of MA training, our findings do not generalize to other forms of math training.

While our results thus far are inconsistent with the view that math training causes improvements to the ANS, they leave open the opposite possibility that ANS acuity might impact mathematics ability. To assess this, our next set of analyses tested whether ANS acuity or estimation performance were related *at all* to math achievement, when other cognitive measures were considered. Critically, the analyses that follow do not hinge on the particularities of the abacus intervention, and instead ask about the relation between the ANS and math independent from math training.

**Relation between math achievement and the ANS.** To test whether ANS acuity was related to math performance across intervention groups, we constructed regression models predicting standardized math scores from ANS acuity. For simplicity, we fit these models for each year separately. These models represent the sort of correlational analyses that are typical of the studies reviewed above that relate ANS acuity to math achievement: They test whether it is possible to use ANS skill to predict concurrent math success.

As in previous research, we found that ANS acuity was a concurrent predictor of standardized math scores for Years 0, 1, and 3 (see Table 2 for *B*, *SE,* and *p*), though we did not observe a significant relationship in Year 2. For these and all subsequently reported models, we scaled all predictors in order to compare the relative predictive value of each parameter in the models directly (since all betas and standard errors are in standard units). ANS was also a concurrent predictor of our math PC1 in Years 0, 2, and 3, but not in Year 1 (Table 2). Thus, while we found inconsistent evidence, the majority of our correlations revealed a concurrent predictive relation between ANS and math, replicating previous results.

**Relation between math achievement and estimation.** We next asked whether estimation performance concurrently predicted math success. Each of our estimation measures concurrently predicted standardized math scores in Years 2 and 3, but not in Year 1 (see Table 2 for statistics). Ordinality was a concurrent predictor of our math PC1 in Years 2 and 3, while both PAE and Linear $r^2$ concurrently predicted our math PC1 every year. Thus, as with previous work, we show strong evidence that estimation performance predicts concurrent math success. Interestingly, it appears as though *both* measures of estimation accuracy and internal consistency were successful in predicting concurrent math outcomes.

Table 2.
*Univariate Models*

| | Standardized Tests | | | PC1 | | |
|---|---|---|---|---|---|---|
| | *B* | *SE* | *p* | *B* | *SE* | *p* |
| Year 0 | | | | | | |
| ANS | -0.286 | 0.076 | 0.0002 | -0.296 | 0.079 | 0.0002 |
| Year 1 | | | | | | |
| ANS | -0.121 | 0.073 | 0.01 | -0.129 | 0.073 | 0.08 |
| PAE | -0.074 | 0.074 | 0.32 | -0.168 | 0.073 | 0.02 |
| Linear R^2 | 0.103 | 0.074 | 0.17 | 0.158 | 0.074 | 0.03 |
| Ordinality | 0.0008 | 0.074 | 0.99 | 0.105 | 0.074 | 0.16 |
| Year 2 | | | | | | |
| ANS | -0.105 | 0.074 | 0.16 | -0.147 | 0.074 | 0.048 |
| PAE | -0.145 | 0.073 | 0.049 | -0.248 | 0.071 | 0.011 |
| Linear R^2 | 0.191 | 0.073 | 0.001 | 0.249 | 0.073 | 0.0008 |
| Ordinality | 0.208 | 0.073 | 0.005 | 0.191 | 0.075 | 0.011 |
| Year 3 | | | | | | |
| ANS | -0.226 | 0.072 | 0.002 | -0.262 | 0.071 | 0.0003 |
| PAE | -0.228 | 0.072 | 0.002 | -0.248 | 0.071 | 0.0007 |
| Linear R^2 | 0.277 | 0.071 | 0.0001 | 0.338 | 0.07 | <.0001 |
| Ordinality | 0.228 | 0.072 | 0.002 | 0.274 | 0.072 | 0.0002 |

[a] Note: each row contains two models; so, there are eight Year 3 models presented in the table.

**Math, ANS, estimation, and other cognitive measures.** Having replicated past work showing that ANS and estimation performance are relatively good concurrent predictors of math performance, we next asked whether these predictive relations still held when controlling for non-numerical tasks. We predicted the same math outcomes (our standardized math score composite and PC1) from our numerical predictors (ANS and estimation performance), while controlling for a host of non-numerical measures (mental rotation, spatial WM, verbal WM, Raven's, age, and Intervention Condition). We constructed one model per year, per dependent variable. This analysis allowed us to ask whether the above-reported correlations between ANS, estimation ability, and math outcomes represent a privileged predictive relation between nonverbal number skill and verbal number competence, or, instead, whether they might be better explained by other, non-numerical predictors. We report the results of our models (standardized betas and SEs) in Figure 1 (predicting standardized test scores) and Figure 2 (predicting PC1). We also experimented with a number of more sophisticated techniques, including longitudinal growth modeling with lagged predictors. Unfortunately, the combination of missing data and the relatively small number of longitudinal data points relative to the large number of possible predictors made these analyses difficult to interpret, though none contradict the results we report here.

While measures like Raven's (Y1-Y3, all models $|B| > .21$), Verbal Working Memory (Y1-Y3, all models $|B| > .09$), Mental Rotation (Y1-Y3, all models $|B| > .11$), and the mental abacus intervention (Y1-Y3, all models $|B| > .16$) predicted standardized math scores with relative consistency (Fig. 1), the ANS and estimation measures did not. For example, when controlling for non-numerical tasks, ANS only significantly predicted standardized test scores in year 0 ($B = -.24$, $SE = .08$, $p = .003$, all other $p>.05$). Linear $r^2$ and PAE never significantly predicted standardized test scores (all $p>.05$; but see Figure 1 for some evidence that Linear $r^2$ may have some predictive power), and Ordinality only predicted standardized test scores in Year 2 ($B = .15$, $SE = .07$, $p = .04$). To summarize, our non-numerical measures, like Raven's and Mental Rotation, consistently predicted standardized test scores, while our numerical ANS and estimation measures only inconsistently contributed to predicting standardized test scores (once we controlled for other factors).

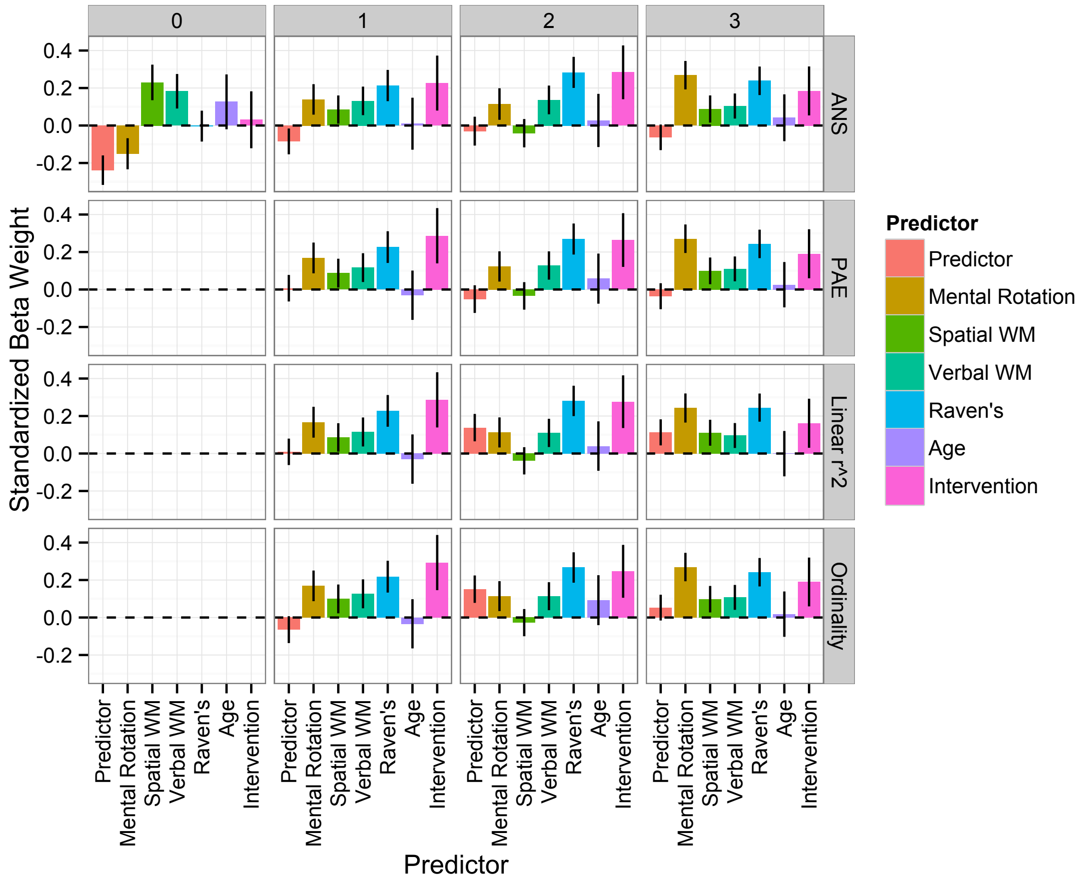


Figure 1. Standardized Beta weights (bars are standard error) when predicting standardized math scores from each of our predictors (ANS, PAE, Linear $r^2$, and Ordinality), controlling for other non-numerical tasks. Each cell represents the results of a single model output, such that the results of 13 models are depicted. Columns represent years of test (Y0-Y3).

Having explored the relationship between standardized math scores and the ANS and estimation tasks, we next applied the same analyses to predict the composite math score (PC1; see Figure 2). Again, measures like Raven's (Y1-Y3, all models $|B| > .19$), Mental Rotation (Y1-Y3, all models $|B| > .08$), the abacus intervention (Y1-Y3, all models $|B| > .19$), and Verbal Working Memory (Y1-Y3, all models $|B| > .07$) all predicted PC1 relatively consistently. Again, ANS performance significantly predicted PC1, but only in Year 0 (*B* = -.19, *SE* = .08, *p*=.02; all other *p*>.05). Neither Ordinality nor PAE ever predicted PC1 when controlling for other factors (all *p*>.05; but see Figure 3 for some evidence that Ordinality may have had predictive value). Interestingly, Linear $r^2$ did significantly predict PC1 in both Year 2 (*B* = .19, *SE* =.07, *p* = .007) and Year 3 (*B* = .17, *SE* = .07, *p* = .008), though not in Year 1 (*p*>.05). Once again, with the exception of Linear $r^2$, our estimation and ANS measures typically failed to predict our math PC1 once other control variables were included in the model.

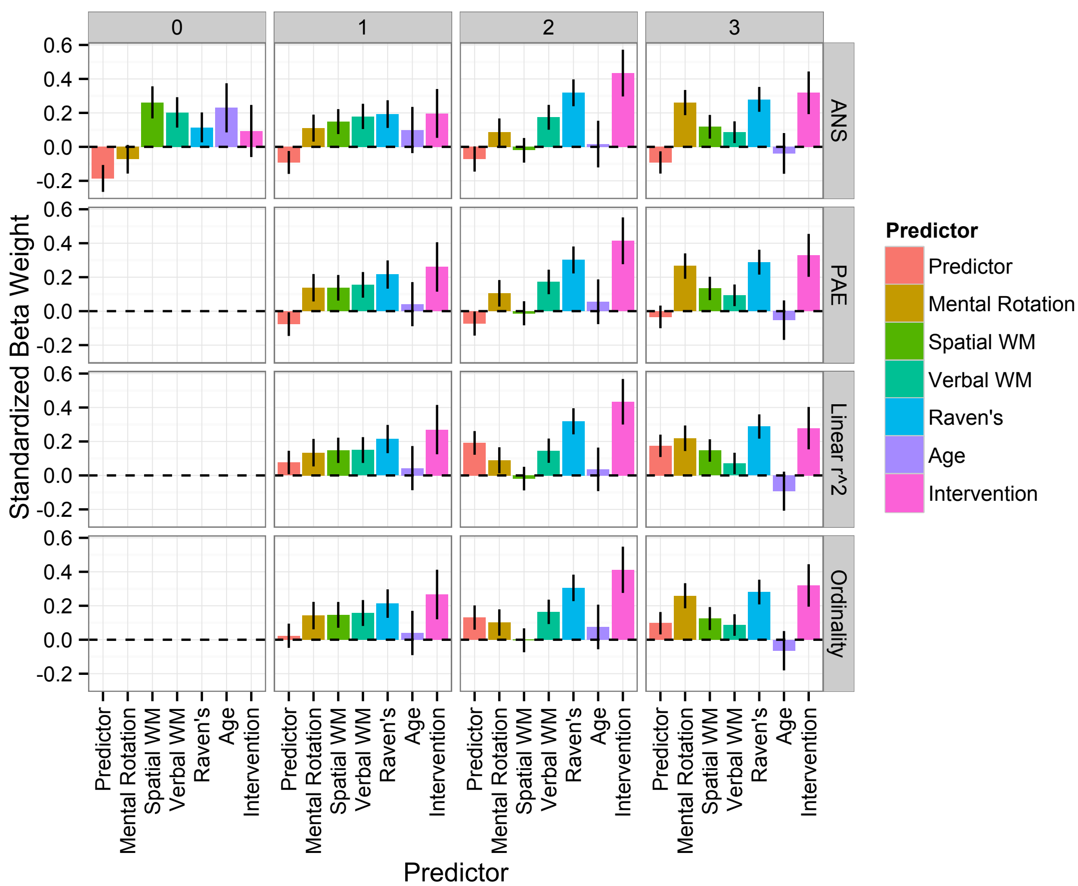

Figure 2. Standardized Beta weights (bars are standard error) when predicting our math PC1 from each of our predictors (ANS, PAE, Linear $r^2$, and Ordinality), controlling for other non-numerical tasks.

## Discussion

In the present study, we tested whether nonverbal number (ANS) acuity and verbal estimation ability were related to symbolic math achievement. New to this study, we assessed (1) whether improvements to math performance caused changes to ANS acuity; (2) whether relations between math performance, ANS, and estimation were consistent over time; and (3) whether relations between math performance, ANS, and estimation persisted over time when controlling for performance on a battery of non-numerical control tasks. To test these questions, we conducted new analyses of data from a 3-year-long randomized controlled math intervention (which led to the intervention group improving substantially at math relative to the control group). This dataset allowed us to test both causal and correlational relations between formal math ability and informal numeracy measures, and to do so in a way that controlled for non-numerical cognitive abilities.

Consistent with previous research, we found that both ANS acuity and estimation performance served as concurrent predictors of math success. However, we also found that this predictive relation was attenuated substantially when other, non-numerical predictors were included in the model. In fact, non-numerical measures like Raven's, Mental Rotation, and verbal working memory were often very strong predictors of math outcome. With the exception of a small subset of our analyses (ANS acuity in Year 0 and Linear $r^2$ in Years 2 and 3), we found little evidence that our ANS and estimation measures uniquely predicted math outcomes, when controlling for other cognitive abilities.

Why might correlations between estimation, ANS acuity, and mathematics achievement disappear when controlling for other cognitive capacities? One likely explanation is that tasks which measure ANS acuity and estimation also depend on capacities like spatial working memory, and domain general abilities like comparison, analogy, and perhaps even proportional reasoning; all of these skills have been implicated in mathematics or estimation performance (Alloway & Passolunghi, 2011; Barth & Paladino, 2011; Clark, Pritchard, & Woodward, 2010; DeStefano & Lefevre, 2004; Geary, 2011; Link, Nuerk, & Moeller, 2013; Thompson, Nuerk, Moeller, & Cohen Kadosh, 2013; Sullivan & Barner, 2014a; b). Because the ANS and estimation tasks that have previously been shown to predict math skill also depend on non-numerical cognitive and perceptual abilities, and because few past studies thoroughly measure these factors, these previous findings may be driven in part by confounding non-numerical factors.

Alternatively, it may be the case that the ability use ANS representations to perform approximate math computations (e.g., the ability to nonverbally "add" quantities) – rather than ANS acuity or magnitude comparison performance – is most predictive of math success. We did not directly test this possibility (and it is somewhat inconsistent with the large literature reporting predictive relations between ANS, estimation, and math performance). However, recent interventions on ANS have shown that, at least in adults, training the use of the ANS during approximate (nonverbal) arithmetic is more effective than the training of simple nonverbal numerical comparisons (like those tested in our study; Park & Brannon, 2013). Thus, it remains possible that there is a privileged relation between the ANS and math outcomes, and our study simply failed to detect it.

Our study had several properties that may limit its generalizability. As with all research in psychology (Henrich et al., 2010), the characteristics of the participants in our sample may limit our ability to draw generalizable inferences about the relation between math, ANS performance, and estimation success in the general human population. Our participants were low-income students whose performance on a variety of tasks was below both Indian and US norms (Barner et al., in press). Thus, our conclusions are tempered by the possibility that predictive relations between approximate and symbolic mathematics might be stronger or different in samples drawn from other populations. Nevertheless, our study suggests that further work on this topic must enroll large, diverse samples and must control for non-numerical cognitive abilities. Also, as already noted, to the extent that the mental abacus intervention failed to affect ANS representations, this may be particular to mental abacus, which is a unique math training program known to recruit non-mathematical cognitive skills and neural regions (Barner et al., in press; Du et al., 2013). Thus, the lack of causal effect of math training on ANS acuity may not be generalizable to children learning mathematics using standard methods.

To conclude, while we replicated past findings that ANS and estimation ability are concurrently predictive of math success, we failed to find evidence that changes to math skill *caused* changes to ANS or estimation performance. We also failed to find consistent evidence that ANS and estimation performance *uniquely* predicted math success. In fact, the strongest predictors of math performance were our non-numerical cognitive predictors, like Raven's, verbal working memory, and mental rotation. These data suggest that, while approximate measures of numerical competence (e.g., ANS acuity and estimation) may be related to math success, their relation in our dataset was fragile, at best.

Acknowledgments. We thank the staff, families, and children at Zenith School for help in collecting data, with a special thank you to Abbasi Barodawala, Mary Joseph, and Snehal Karia. Thanks also to George Alvarez, for stimulus design and funding support. Thanks also to Neon Brooks, Sean Barner, Eleanor Chestnut, Jonathan Gill, Ali Horowitz, Talia Konkle, Ally Kraus, Molly Lewis, Bria Long, Ann Nordmeyer, Mahesh Srinivasan, Viola Störmer, Jordan Suchow, and Katharine Tillman for help with data collection. This work was funded by a grant to D.B. and George Alvarez. from NSF REESE grant #0910206 and by an NSF GRFP to J.S.

Alloway, T. & Passolunghi, M., (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences, 21,* 133 - 137. doi: 10.1016/j.lindiff.2010.09.013

Anobile, G., Stievano, P., & Burr, D. (2013). Visual sustained attention and numerosity sensitivity correlate with math achievement in children. *Journal of Experimental Child Psychology, 116,* 380-391.

Barner, D., Alvarez, G., Sullivan, J., Brooks, N., Srinivasan, M., & Frank, M. (in press). Learning mathematics in visuospatial format: A randomized, controlled trial of mental abacus instruction. *Child Development*.

Barth, H. & Paladino, A. (2011). The development of numerical estimation in children: evidence against a representational shift. *Developmental Science, 14,* 125-135. doi: 10.1111/j.1467-7687.2010.00962.x

Bonny, J. & Lourenco, S. (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. doi: 10.1016/j.jecp.2012.09.015

Booth, J., & Siegler, R. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 41,* 189-201. doi: 10.1037/0012-1649.41.6.189.

Booth, J., & Siegler, R. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79,* 1016-1031. DOI: 10.1111/j.1467 - 8624.2008.01173.x

Castronovo, J., & Göbel, S. (2012). Impact of high mathematics education on the number sense. *PLoS ONE, 7,* e33832. doi: 10.1037/journal/pone.0033832.

Clark, C., Pritchard, V., & Woodward, L. (2010). Preschool executive function abilities predict early mathematics achievement. *Developmental Psychology, 46,* 1176 - 1191. doi: 10.1037/a0019672

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.

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press: New York.

Dehaene, S., Izard, V., & Piazza, M. (2005). Control over non-numerical parameters in numerosity experiments. *Unpublished manuscript (available on www.unicog.org).*

De Smedt, B., Noel, M., 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.

Desoete, A., Ceulemans, A., De Weerdt, F., & Pieters, S. (2010). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparisons tasks in kindergarten? Findings from a longitudinal study. *British Journal of Educational Psychology, 82,* 64-81.

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16,* 353-386. doi: 10.1080/09541440244000328.

DeWind, N., & Brannon, E. (2012). Malleability of the approximate number system: effects of feedback and training. *Frontiers in Human Neuroscience, 6,* 68. doi:

10.3389/fnhum.2012.00068.

Du, F., Chen, F., Li, Y., Hu, Y., Tian, M., & Zhang, H. (2013). Abacus training modulates the neural correlates of exact and approximate calculates in chinese children: An fMRI study. *BioMed Research International, 2013,* 1-12. doi: 10.1155/2013/694075.

Frank, M., & Barner, D. (2011). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, *141,* 134-149. DOI: 10.1037/a0024427

Fuhs, M., & McNeil, N., (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Developmental Science, 16,* 136-148.

Geary, D. (2011). Cognitive predictors of achievement growth in mathematics: A five year longitudinal study. *Developmental Psychology, 47,* 1539-1152. doi: 10.1037/a0025510.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS One, 8,* e67374. 10.1371/journal.pone.0067374.

Gilmore, C., McCarthy, S., & Spelke, E. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115,* 394-406.

Göbel, S., Watson, S., Lervag, A., & Hulme, C. (2014). Children's Arithmetic Development: It is Number Knowledge, Not the Approximate Number Sense, that

Counts. *Psychological Science, 25,* 789-798.

Gunderson, E., Ramirez, G., Beilock, S., & Levine, S. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology, 48,* 1229-1241. doi: 10.1037/a0028593.

Halberda, J., Ly, R., Wilmer, J., Naiman, D., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Science, 109,* 11116-11120. doi: 10.1073/pnas.1200196109

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

Hatano, G., Miyake, Y. & Binks, M. G. (1977). Performance of expert abacus operators. *Cognition*, *5*, 57-71.

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 61-83.

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

Hornung, C., Schlitz, C,. Brunner, M., & Martin, R. (2014). Predicting first-grade mathematics achievement: the contribution of domain-general cognitive abilities, nonverbal number sense, and early number competence. *Frontiers in Psychology, 5,* 272.

Kolkman, M., Kroesbergen, E., & Leseman, P. (2013). Early numerical development and the role of non-symbolic and symbolic number skills. *Learning and Instruction, 25,* 95-103.

Libertus, M., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14,* 1292-1300.

Libertus, M., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences, 25,* 126 -133. doi: 10.1016/j.lindif.2013.02.001

Libertus, M., Odic, D., Feigenson, L., & Halberda, J. (2015). Verbal number estimation predicts math ability and mediates the relation between numerical approximation and math ability. *Society for Research on Child Development, Philadelphia, PA.*

Libertus, M., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141,* 373-379. doi: 10.1016/j.actpsy.2012.09.009

Lindskog, M., Winman, A., & Juslin, P. (2014). Association between higher education and approximate number system acuity. *Frontiers in Psychology, 5,* 462.

Link, T., Nuerk, H.C., & Moeller, K., (2014). On the relation between the mental number -line and arithmetic competence. *The Quarterly Journal of Experimental Psychology, 67,*1597-1613. doi: 10.1080/17470218.2014.892517

Lourenco, S., Bonny, J., Fernandez, E., & 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, 109*, 18737-18742. doi: 10.1073/pnas.1207212109

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

Lyons, I., Price, G., Anniek,V., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science, 17,* 714-726.

Mazzocco, M., Feigenson, L., & Halberda, J. (2011a). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS Oone*, *6*, e23749. doi: 10.1371/journal.pone.0023749

Mazzocco, M., Feigenson, L., & Halberda, J. (2011b). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*, 1224-1237. doi:10.1111/j.1467 -8624.2011.01608.x

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLOSOne, 8,* e67918.

Moore, A., & Ashcraft, M. (2015). Children's mathematical performance: Five cognitive tasks across five grades. *Journal of Experimental Child Psychology, 135,* 1-24. doi: 10.1016/j.jecp.2015.02.003.

Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., & Content, A. (2013). Does math education modify the approximate number system? A comparison of schooled and unschooled adults. *Trends in Neuroscience and Education, 2,* 13-22.

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction, 23,* 125-135.

Park, J. & Brannon, E. (2013). Training the Approximate Number System Improves Math Proficiency. *Psychological Science, 24,* 2013-2019. doi: 10.1177/0956797613482944

Park, J., & Brannon, E. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition, 133,* 188-200. doi: 10.1016/j.cognition.2014.06.011.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S. & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*, 33 - 41. doi:10.1016/j.cognition.2010.03.012

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science, 306,* 499-503.

Price, G., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica, 140,* 50-57. doi: 10.1016/j.actpsy.2012.02.008.

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *The British Journal of*

*Developmental Psychology, 30,* 344-357.

Sasanguie, D., Göbel, S., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114,* 418-431.

Sasanguie, E., Defever, B., 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. doi: 10.1080/17470218.2013.803581

Siegler, R., & Booth, J. (2004). Development of numerical estimation in young children. *Child Development, 75,* 428-444. DOI: 10.1111/j.1467-8624.2004.00684.x

Starr, A., Libertus, M., & Brannon, E. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences, 110,* 18116-18120. doi: 10.1073/pnas.1302751110

Stigler, J. W., Chalip, L., & Miller, K. (1986). Consequences of skill: The case of abacus training in Taiwan. *American Journal of Education*, *94*(4), 447–479.

Sullivan, J., & Barner, D. (2014a). Inference and association in children's early numerical estimation. *Child Development, 85*(4), 1740- 1755.

Sullivan, J., Barner, D. (2014b). The development of structural analogy in number-line estimation. *Journal of Experimental Child Psychology*, 128, 171-189. doi: 10.1016/j.jecp.2014.07.004

Tibber, M., Manasseh, G., Clarke, R., Gagin, G., Swanbeck, S., Butterworth, B., Lotto, R., & Dakin, S. (2013). Sensitivity to numerosity is not a unique visuospatial psychophysical predictor of mathematical ability. *Vision Research, 89,* 1-9. doi:10.1016/jvisres.2013.06.006.

Thompson, J., Nuerk, H., Moeller, K., & Cohen Kadosh, R. (2013). The link between mental rotation ability and basic numerical representations. *Acta Psychologica, 144,* 324-331. doi: 10.1016/j.actpsy.2013.05.009

Wei, W., Yuan, H., Chen, C., & Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *British Journal of Educational Psychology, 82,* 157-181.

Xu, F., & Spelke, E. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74,* B1-B11. doi: 10.1016/S0010-0277(99)00066-9