

Research Reports

Intensive Math Training Does not Affect Approximate Number Acuity: Evidence From a Three-Year Longitudinal Curriculum InterventionJessica Sullivan^{*a}, Michael C. Frank^b, David Barner^{cd}

[a] Department of Psychology, Skidmore College, Saratoga Springs, NY, USA. [b] Department of Psychology, Stanford University, Stanford, CA, USA. [c] Department of Psychology, University of California, San Diego, CA, USA. [d] Department of Linguistics, University of California, San Diego, CA, USA.

Abstract

Does nonverbal, approximate number acuity predict mathematics performance? Some studies report a correlation between acuity of representations in the Approximate Number System (ANS) and early math achievement, while others do not. Few previous reports have addressed (1) whether reported correlations remain when other domain-general capacities are considered, and (2) whether such correlations are causal. In the present study, we addressed both questions using a large ($N = 204$) 3-year longitudinal dataset from a successful math intervention, which included a wide array of non-numerical cognitive tasks. While we replicated past work finding correlations between approximate number acuity and math success, these correlations were very small when other domain-general capacities were considered. Also, we found no evidence that changes to math performance induced changes to approximate number acuity, militating against one class of causal accounts.

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

Journal of Numerical Cognition, 2016, Vol. 2(2), 57–76, doi:10.5964/jnc.v2i2.19

Received: 2015-12-10. Accepted: 2016-04-04. Published (VoR): 2016-08-05.

*Corresponding author at: Department of Psychology, Skidmore College, 815 N. Broadway, Saratoga Springs, NY, 12866. Phone: (518)-580-5310; Fax: 518-580-5319. E-mail: jsulliv1@skidmore.edu



This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Beginning early in infancy, humans can represent approximate numerical quantities nonverbally, using what is sometimes called the “Approximate Number System” (ANS) or the “number sense” (Dehaene, 1997; Xu & Spelke, 2000). The ANS is used to represent and compare numerical magnitudes, and does so according to Weber’s law, such that the ratio of any two numerical quantities determines the likelihood that 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 (Chen & Li, 2014; Halberda, Mazocco, & Feigenson, 2008; Halberda et al., 2012).

In fact, more than a dozen studies have reported some correlation between the ANS and symbolic math, and these correlations often survive 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, Mazocco, & Feigenson, 2008; Halberda et al., 2012; Libertus et al., 2011; Libertus, Feigenson, & Halberda, 2013; Libertus, Odic, & Halberda, 2012; Mazocco et al., 2011a,

2011b; 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. 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.

Many other studies, however, 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, knowledge of numerical cardinality, and non-numerical quantity comparison (e.g., Chu et al., 2015; 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 et al., 2013; Sasanguie, Defever, Maertens, & Reynvoet, 2014; Sasanguie, De Smedt, Defever, & Reynvoet, 2012; Tibber et al., 2013; van Marle, Chu, Li, & Geary, 2014; Wei et al., 2012). Further, at least one study that trained children's ANS acuity found no effect on math ability (Obersteiner, Reiss, & Ufer, 2013). These discrepant findings raise important questions about the nature (and malleability) of the ANS, and about the practical significance of any relationship between the ANS and symbolic math (for review, see De Smedt, Noël, 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 (36 independent) samples – ANS acuity explained substantial variability in symbolic math achievement (Chen & Li, 2014). However, 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 controlled only for participants' linguistic ability. This inconsistent inclusion of control tasks is a problem, because tasks typically used to measure ANS acuity – e.g., dot array comparison – could plausibly draw on other non-numerical cognitive capacities, like working memory, non-numerical quantity representation, and inhibitory control (e.g., Gilmore et al., 2013). Critically, each of these cognitive capacities has also been shown to predict early math achievement, making it possible that they mediate relationship between ANS acuity and math achievement (Alloway & Passolunghi, 2011; Clark, Pritchard, & Woodward, 2010; DeStefano & LeFevre, 2004; Geary, 2011; Gilmore et al., 2013; Hornung, Schiltz, Brunner, & Martin, 2014; Lourenco, Bonny, Fernandez, & Rao, 2012; Passolunghi, Cargnelutti, & Pastore, 2014; Thompson, Nuerk, Moeller, & Cohen Kadosh, 2013). Thus, although the 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. 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; Nys et al., 2013; Pica, Lemer, Izard, & Dehaene, 2004).

In the present study, we tested whether the ANS is meaningfully causally linked to mathematics achievement. We did this first, by assessing its longitudinal predictive power relative to a large battery of other cognitive measures, and second, by assessing whether changes in math performance *caused* changes to the ANS. Specifically, we analyzed publicly available data from a three-year longitudinal randomized controlled trial (RCT) of a math intervention in 2nd through 5th graders ($N = 204$). This RCT provided supplemental mathematics training in an experimental group using a popular mental arithmetic technique called “mental abacus” (Barner et al., 2016). Mental

abacus training involves teaching participants to perform arithmetic calculations using an 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, 2012; Hatano, Miyake, & Binks, 1977; Stigler, Chalip, & Miller, 1986). In the RCT, abacus training improved children's arithmetic abilities relative to a control group who had received additional training in a standard math curriculum, allowing us to test whether changes in arithmetic skill (induced by the intervention) led to changes in ANS acuity. Critically, the study also included a longitudinally-administered measure of ANS acuity, as well as tests of spatial working memory, verbal working memory, mental rotation ability, and general intelligence (Raven's Progressive Matrices). The availability of these measures allowed us to assess the predictive relation between ANS acuity and math achievement while concurrently controlling for a larger than usual set of non-numerical cognitive capacities. To our knowledge, this is the first study to have experimentally manipulated math ability in order to test the effects of improvements to math skill on ANS acuity.

In addition to the measures described above, Barner et al. (2016) also collected (but did not analyze) longitudinal dot-array estimation data, by measuring children's ability to label arrays of dots with number words. These data allowed us to test the relationship between estimation and math achievement. Given that ANS representations are known to be linked to number words in the verbal count list (e.g., Le Corre & Carey, 2007; Mundy & Gilmore, 2009; Whalen, Gallistel, & Gelman, 1999), 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 (Booth & Siegler, 2006, 2008; Gunderson, Ramirez, Beilock, & Levine, 2012; Kolkman et al., 2013; Moore & Ashcraft, 2015; Siegler & Booth, 2004). To our knowledge, no previous study has assessed this link while simultaneously controlling for a large battery of domain-general cognitive capacities.

To summarize, we re-analyzed data from a longitudinal math intervention to test the uniqueness and causal status of correlations between the ANS and math achievement, while simultaneously probing dot-array estimation, a measure of associations between numerals and approximate magnitudes. In doing so, we provide the first large-scale longitudinal study to assess the causal link between the ANS and math achievement while simultaneously controlling for an exhaustive battery of domain-general cognitive capacities.

Method

Participants

Data were obtained from a previous study by Barner et al. (2016) 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 (Gujarati 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 (median household income for 2006-2012 was just over \$9,700 globally, and just over \$3,000 in India; Phelps & Crabtree, 2013). Children were between of 5 and 8 years of age ($M = 6.65$ years, $SD = .53$) at the time of enrollment in 2010 (Year 0).

Math Measures

Children received several measures of math competence, including the Woodcock-Johnson III Computation test, the Math Fluency subtest of the Wechsler Individual Achievement Test (WIAT-III), and in-house tests of arithmetic and place value understanding. Also, children's math grades were available (as a score between 0-100), as reported by their school. Detailed descriptions of these measures are available in the Supplementary Materials of Barner et al. (2016).

Intervention

As reported in Barner et al. (2016), all children followed a standard math curriculum throughout the 3 years of the study. In addition to their standard math curriculum, students were also enrolled in one of two *supplemental* math programs: The control program used a standard international math curriculum, while the abacus program taught students mental abacus. Thus, all children received a shared core math curriculum, and one of two possible supplemental math curricula. The random assignment procedure operated as follows. At enrollment in Year 0, children were randomly assigned to one of three homerooms. Children randomly assigned to Homeroom 1 received supplemental instruction in a standard international math curriculum. Children assigned to Homeroom 2 received an equal number of hours of supplemental training using mental abacus. Half of the children in Homeroom 3 received supplemental abacus training, and the other half received the control supplemental standard math curriculum. Supplemental abacus instruction was conducted by a privately trained abacus instructor, and in the case of the split class was conducted in a space separate from control children. Of the 204 children in the study, 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 related to changes in 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 timed computerized task. As is typical for tasks assessing ANS acuity, 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 participants 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). Ratios ranged from 1:2 to 15:16.

Children's estimation ability was tested by asking them to estimate the number of dots on a screen; task duration was 10 minutes. 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 (errors could be corrected by using the backspace key). Prior to beginning the task, children completed a keypad typing training session to ensure that all children could appropriately use the keypad.

Control Tasks

Children were tested on a battery of control tasks, again described in detail in the Supplemental Materials of [Barner et al. \(2016\)](#). This battery included two computerized measures of working memory: (1) a test of verbal working memory; and (2) a test of spatial working memory. For both tasks, the participant was first presented with a target sequence of events, and then had to decide whether a second sequence of events was the same as or different from the target (see below). For both tasks, children completed 10 practice trials prior to the start of the experiment, and 24 trials as part of the experiment. Children provided their responses by pressing either the 's' key (reabeled as "S" for same) or the 'l' key (reabeled as "D" for different) on a keyboard.

In the verbal working memory task, children heard a sequence of target syllables (e.g., "GU, TI"). They then had to decide whether a second sequence of syllables (e.g., "RA", "TI") was the same or different from the target. The test sequence was either identical to the target, or differed by one syllable. All stimuli were recorded by a native speaker of Gujarati, using syllables that are legal in English, Hindi, and Gujarati.

In the spatial working memory task, participants saw a 5x5 grid of circles. They then saw a target sequence of circles turn yellow (only one circle was filled at a time). They were then presented with a test sequence of blue circles, and had to decide whether the blue circles appeared in the same locations as the yellow circles. Again, the sequence of blue circles was either identical to the sequence of yellow circles, or contained a single circle that appeared in a different position.

Both working memory tasks were adaptive – children were asked to remember n items during the target sequence, and if they responded correctly, they were asked on the next trial to remember $n+1$ items (if they responded incorrectly, they were next asked to remember $n-1$ items). Thus, if a child successfully remembered two target syllables (e.g., "GU", "TI"), on the next trial they were asked to remember three target syllables (e.g., "RA", "MI", "TU"). For both working memory tasks, we calculated participants' memory score by averaging the level of difficulty for all 24 trials.

Children also completed a paper-and-pencil task that measured mental rotation ability. In this task, participants were asked to match one of two target items (either letters or shapes, based on the [Shepard & Metzler, 1971](#) stimuli) to a sample. One target item was the mirror image of the sample while the other was an exact match; both target items were also rotated. Thus, in order to determine which target matched the sample, the child needed to mentally rotate the item. Finally, children completed Raven's Progressive Matrices (A and/or B).

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 measured for each year (Year 0, 1, 2, and 3) as a Weber fraction (using the method described by Halberda, Mazocco, & Feigenson, 2008), where greater ANS acuity is indicated by a smaller Weber fraction (M_{acuity} Y0 = .37; Y1 = .18; Y2 = .15; Y3 = .14). Prior to analysis, we excluded children whose Weber fractions were $> .8$, as a value this large likely reflects a misunderstanding of the task. We selected the Weber fraction as our DV because it is the standard measure of ANS acuity in the field. However, because measures of ANS acuity differ from one another (Inglis & Gilmore, 2014), we also repeated all analyses using an alternative, less conservative measure of acuity, defined as the ratio of dots on the most difficult trial successfully completed by a child. These additional analyses are available here: <https://github.com/langcog/jesestimation>, and are only reported in the present paper when they differed from our main analyses.

Using previously unreported data, 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. Specifically, we defined ordinality 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 was 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, demonstrating high levels of ordinal responding. Surprisingly, performance on this measure did not appear to improve over time. 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; again, these values did not appear to increase over time). 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 (Booth & Siegler, 2006, 2008; Gunderson, Ramirez, Beilock, & Levine, 2012; Kolkman et al., 2013; Moore & Ashcraft, 2015; Siegler & Booth, 2004).

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_{PAE} : $Y1 = .71$; $Y2 = .71$; $Y3 = .70$; again, no group-wise change over time). PAE has been shown previously to predict math performance on standardized tests (Castronovo & Göbel, 2012; Sasanguie et al., 2013; Siegler & Booth, 2004; although not across all studies: see Booth & Siegler, 2006), addition/subtraction performance (Link, Nuerk, & 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.

Year	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***

* $p < .05$. ** $p < .01$. *** $p < .001$.

Math Measures

We did not have a set of specific, a priori, theoretically-motivated predictions about the differences between our particular measures of math competence (WIAT, WJ-III, arithmetic, place value, and math grades), and so we constructed two different math measures, both of which combined data from multiple math tests. 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.

We created a composite of the standardized math test scores (WIAT math fluency and WJ-III Computation subtest) by calculating the proportion correct for each test, and averaging scores on the two measures. This composite showed improved mathematics performance across each year of testing ($M_{\text{standardized}}$: $Y0 = .20$; $Y1 = .31$; $Y2 = .43$; $Y3 = .54$). Because standardized math testing is commonly used both in psychology and education, this measure captured the type of math competence that is likely to be measured in a classroom or lab setting.

We also created a single composite measure that took into account *all* available math achievement data for each child. To do this, 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.

Analyses

Effect of Intervention on ANS and Estimation

Barner et al. (2016), reported using the current data that abacus training had a significant impact on mathematics achievement when measured by the WJ-III Computation subtest and the in-house arithmetic battery. Therefore, by comparing the ANS acuity and estimation ability of children who received abacus instruction to that of children in the control group, we were able to ask whether improvements to mathematics ability caused improvements to ANS and estimation outcomes. 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. If improvements to math ability cause improvements to ANS acuity, then we should expect 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 older children's smaller Weber fractions). However, we found no effect of Intervention Condition ($B = -.02$, $SE = .01$, $p = .11$) or interaction of Intervention Condition and Year ($B = .006$, $SE = .007$, $p = .36$), and planned t -tests also revealed no effects of abacus training on ANS in Years 1 ($t(179) = -.20$, $p = .84$, $d = -.03$), 2 ($t(183) = -.34$, $p = .74$, $d = -.05$), or 3 ($t(183) = -1.26$, $p = .26$, $d = -.19$; Fig. 1). When using our alternate measure of ANS acuity (the hardest ratio reached), the abacus and control group differed somewhat in ANS acuity in Years 1 ($t(182) = 2.29$, $p = .023$, $d = .34$; Bonferroni $p = .068$) and 3 ($t(183) = 2.37$, $p = .019$, $d = .35$; Bonferroni $p = .056$), but these differences failed to reach significance when correcting for multiple comparisons.

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 ($t(182) = -1.64$, $p = .10$, $d = -.24$), a significant, uncorrected, effect in Year 2 ($t(185) = -2.21$, $p = .03$, $d = -.32$; Bonferroni $p = .09$), and no effect in Year 3 ($t(184) = -1.47$, $p = .14$, $d = -.22$). For Ordinality, there was a significant, uncorrected, effect in Year 1 ($t(182) = 2.15$, $p = .03$, $d = .32$; Bonferroni $p = .11$), and no effect in Years 2 or 3 (year 2: $t(185) = 1.39$, $p = .17$, $d = .20$; year 3: $t(184) = 1.12$, $p = .26$, $d = .17$). For Linear r^2 , there were no effect in Years 1 or 2 (Year 1: $t(182) = .598$, $p = .55$, $d = .09$; Year 2: $t(185) = .38$, $p = .71$, $d = .06$), though there was a significant, uncorrected, effect in Year 3 ($t(184) = 2.34$, $p = .02$, $d = .35$; Bonferroni $p = .06$). Although there were trends indicating a relation between Intervention Condition and estimation performance (e.g., see Figure 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, analyses 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.

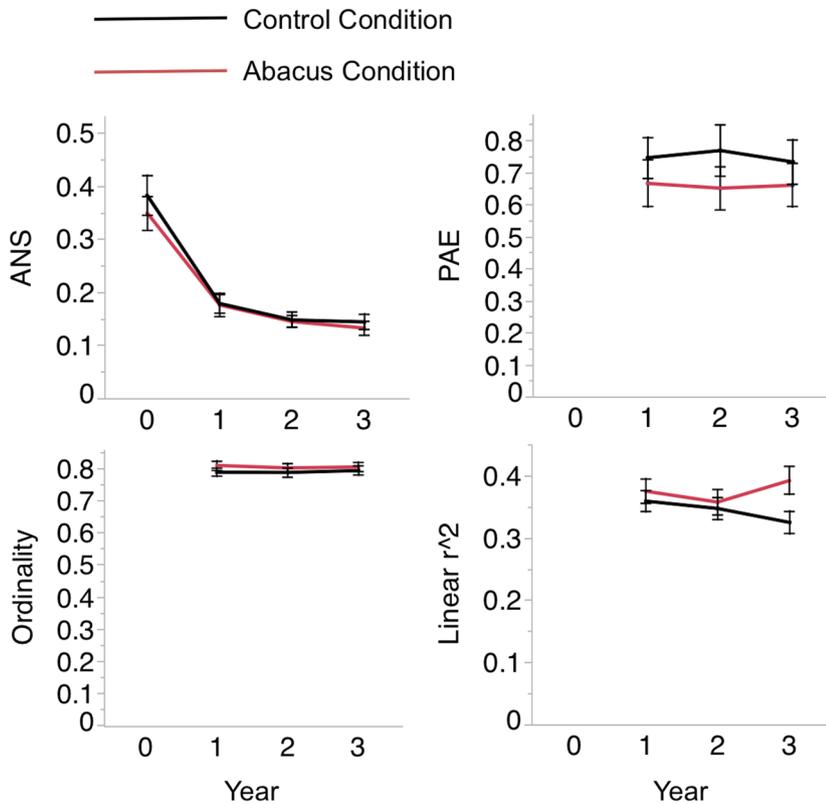


Figure 1. ANS acuity (Years 0, 1, 2, and 3) and estimation performance (Years 1, 2, and 3 – we did not test estimation in Year 0). Red line indicates children who learned abacus; black line indicates children who were in the control group. For Proportion Absolute Error (PAE) and ANS measures, smaller numbers indicate better performance. For our Linear r^2 and Ordinality measures, larger numbers indicate better performance. Error bars are Standard Errors.

While our results thus far are not consistent with the view that math training causes improvements to the ANS, they leave open the opposite possibility that ANS acuity might still be related to mathematics ability. To assess this, our next set of analyses tested whether ANS acuity and 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. Following the logic of previous studies in this literature, these models test whether ANS acuity predicts concurrent math achievement. 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).

Replicating 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. ANS was also a concurrent predictor of our math PC1 in Years 0, 2, and 3, but not in Year 1 (Table 2).¹ Our

alternative measure of ANS acuity (hardest ratio reached) showed that ANS was a concurrent predictor of PC1 in all years. Thus, while we found inconsistent evidence, the majority of our correlations revealed a concurrent predictive relation between ANS and math, replicating previous results.

Table 2

Univariate Models

Predictor	Standardized Tests			PC1		
	<i>B</i>	<i>SE</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>p</i>
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

Note. Each row contains two models so there are (for example) eight Year 3 models presented in the table.

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 in previous work, we find that estimation performance predicts concurrent math achievement. More interesting, however, is whether this relationship survives the addition of a large battery of domain general measures.

Relation Between Domain-General Cognitive Mechanisms, ANS, Estimation, and Math

Having replicated past work showing that ANS and estimation ability are predictive of concurrent math achievement, we next asked whether such predictive relations were uniquely numerical, or whether they could be explained by domain-general cognitive abilities. To do this, we predicted math outcomes (our standardized math score composite and PC1) from numerical predictors (ANS and estimation performance) and from our battery of non-numerical measures (mental rotation, spatial WM, verbal WM, Raven's, age, and Intervention Condition; additional information about each of these measures is available here: <https://github.com/langcog/mentalabacus>). We constructed one model per year, per dependent variable. We report the results of our models (standardized betas and SEs) in

Figure 2 (predicting standardized test scores) and Figure 3 (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.

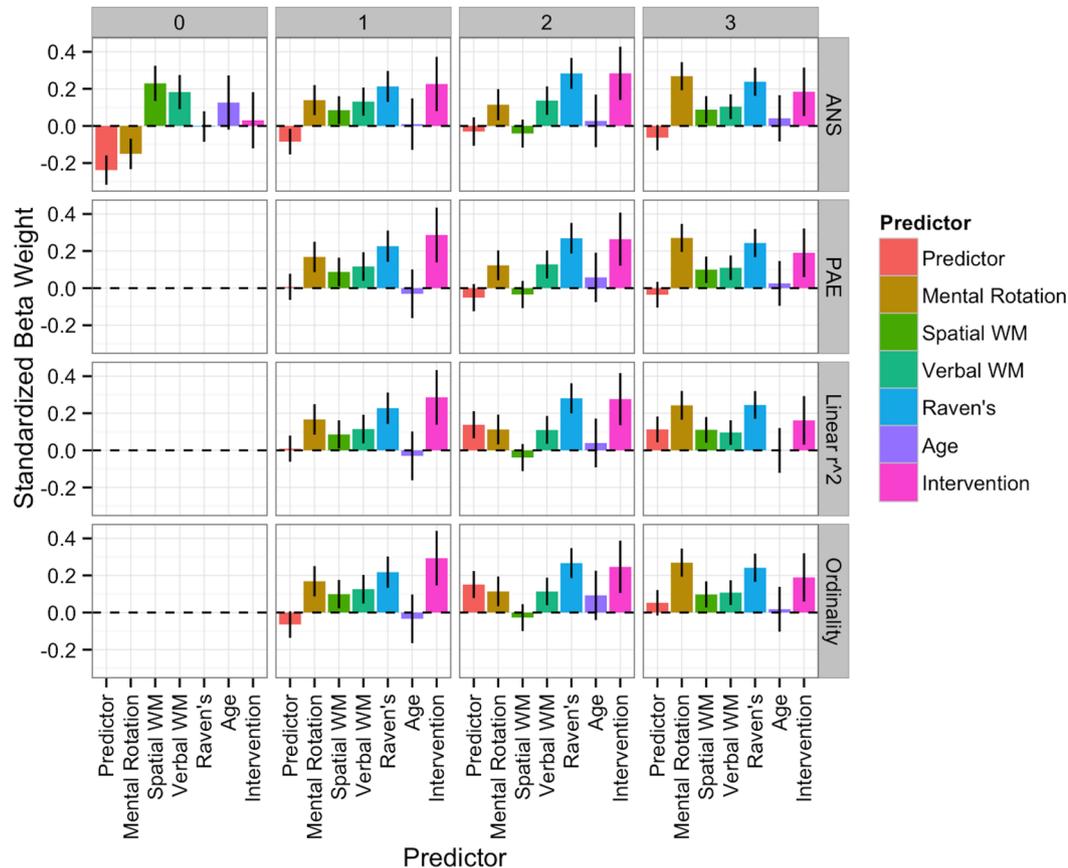


Figure 2. Standardized Beta weights (bars are standard error) when predicting standardized math scores from each of our predictors (ANS, PAE, Linear r^2 , and Ordinality), controlling for other non-numerical tasks. Each cell represents the results of a single model output, such that the results of 13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

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 $|B| > .16$) all predicted standardized math scores with relative consistency (Figure 1), the ANS and estimation measures did not. For example, when controlling for non-numerical tasks, ANS only 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 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).

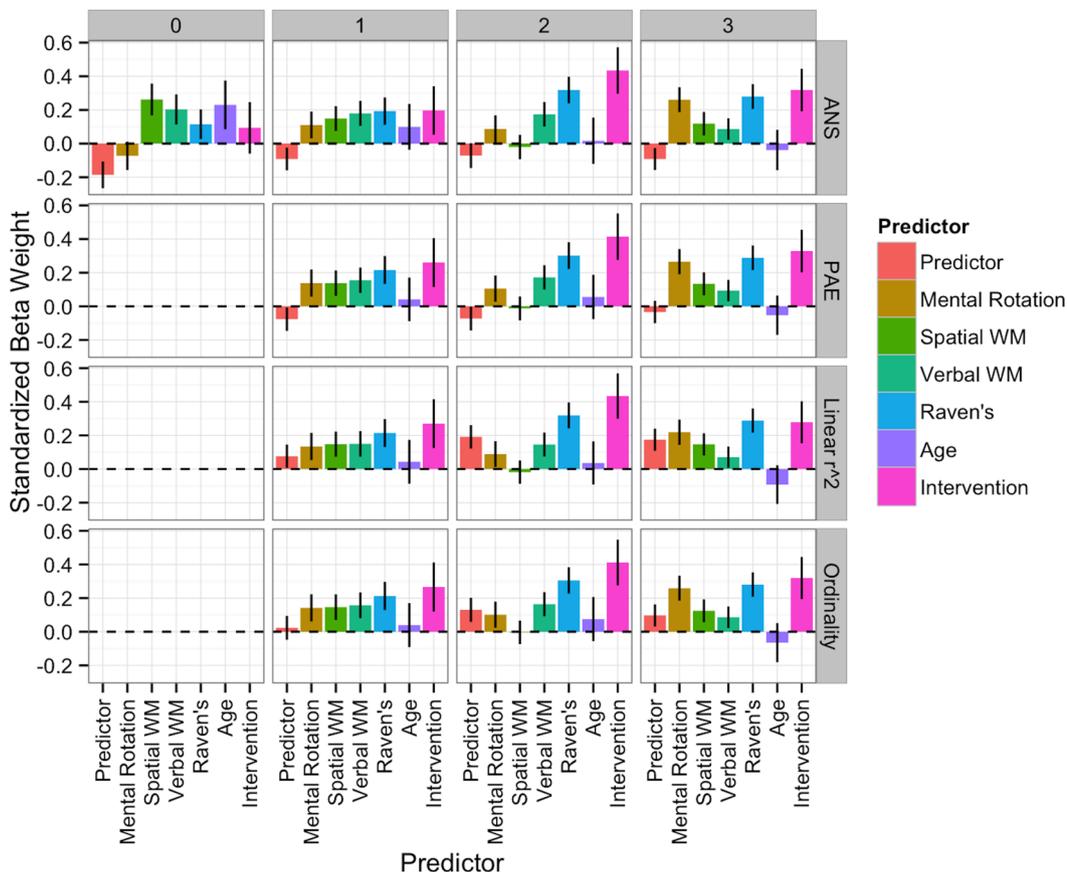


Figure 3. Standardized Beta weights (bars are standard error) when predicting our math PC1 from each of our predictors (ANS, PAE, Linear r^2 , and Ordinality), controlling for other non-numerical tasks.

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 $p > .05$).ⁱⁱ 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 measures typically failed to predict our math PC1 once other control variables were included in the model. Further, our ANS measure did not consistently predict math PC1 when controlling for other factors.

Discussion

We tested whether nonverbal number (ANS) acuity and verbal estimation ability were uniquely predictive of symbolic math achievement. Specifically, 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 three-year-long randomized controlled math intervention, [Barner et al. \(2016\)](#) having reported that this intervention improved math substantially relative to the control group. Unlike any previous study, this dataset allowed us here to test both correlations and possible causal relations between formal math ability and informal numeracy measures, and to do so in a way that robustly controlled for non-numerical cognitive abilities. While previous studies have reported conflicting results using a host of less robust designs, we both replicated positive findings and showed that – in our data – these results are explained almost entirely by domain-general factors.

We first asked whether changes to math performance *caused* longitudinal changes to ANS acuity. While previous studies tested this possibility indirectly (e.g., by comparing ANS acuity between educated and uneducated participants; [Pica et al., 2004](#)), or tested the opposite causal direction (whether improving ANS performance improves math; [Park & Brannon, 2013, 2014](#)) our study was unique in that it asked whether a RCT math intervention (which improved math performance) affected participants' ANS acuity. We found no compelling evidence that this successful math intervention improved ANS acuity, or led to differences in estimation performance. These findings militate against one class of possible causal relations between the ANS and mathematics achievement – e.g., whereby changes in math achievement cause changes that result in greater ANS acuity. One limitation of our conclusions, however, is that mental abacus differs in many ways from other methods for improving math performance: mental abacus training appears to recruit cognitive skills not directly related to math ([Barner et al., 2016](#); [Frank & Barner, 2012](#)), and abacus processing activates *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 mental abacus training, our findings do not generalize to other forms of math training.

Next, we sought to replicate and then explain previous work which found that ANS and estimation ability are related to formal math skill. Consistent with some previous studies, 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 very strong predictors of math outcome, whereas ANS acuity and estimation were not. 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. This finding supports the conclusion that the relation between ANS acuity and symbolic math performance is often weakest in the early elementary school years ([Fazio et al., 2014](#)), and with work showing that this relationship may be mediated by other, non-ANS-related, factors (e.g., [Gilmore et al., 2013](#); [Göbel et al., 2014](#); [Holloway & Ansari, 2009](#)).

Why might correlations between estimation, ANS acuity, and mathematics achievement disappear when controlling for other cognitive capacities? One likely explanation is that tasks that 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, 2014](#); [Passolunghi et al., 2014](#); [Sullivan & Barner, 2014a, 2014b](#); [Thompson, Nuerk, Moeller, & Cohen Kadosh, 2013](#)). Similarly, the visual demands of most ANS tasks

(e.g., Clayton et al., 2015) leave open the possibility that visual processing ability might partially mediate the link between ANS acuity and math performance.ⁱⁱⁱ Because the ANS and estimation tasks that have previously been shown to predict math skill also depend on non-numerical cognitive abilities, and because few past studies thoroughly measured these factors, these previous findings may be driven in part by confounding non-numerical factors.

Alternatively, some have argued that the ability to use ANS representations to perform approximate math computations (e.g., the ability to nonverbally “add” quantities) – rather than ANS acuity itself – is most predictive of math achievement (Park & Brannon, 2013). Consistent with this, recent interventions on ANS acuity have shown that, at least in adults, training the use of the ANS during approximate (nonverbal) arithmetic is more effective than training simple nonverbal numerical comparisons (like those tested in our study; Park & Brannon, 2013). These findings raise the possibility that there is a privileged relation between the ANS and math outcomes, and that our study simply failed to detect it. However, our results raise several additional questions. In particular, our intervention results suggest that improvements to ANS acuity are not required in order for improvements in math performance to occur, thus leaving open the question of why researchers so frequently find correlations between ANS acuity and math performance. Second, the mechanism by which practicing approximate addition might affect symbolic math remains unclear.

Before concluding, we note that while our study serves as an important extension of existing work on the ANS, it remains possible that studies in other populations will yield different results. As in all psychological research, the characteristics of the participants in our sample may limit our ability to draw generalizable inferences in the general human population (Henrich et al., 2010). Our participants were low-income students whose performance on a variety of tasks was below both Indian and US norms (Barner et al., 2016). Thus, the predictive relations between approximate and symbolic mathematics may be different in samples drawn from other populations. Finally, 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., 2016; 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, this relationship is likely fragile, and is one among many that predict mathematics achievement. More informative predictors of math achievement include domain general capacities like working memory, mental rotation, and general intelligence.

Notes

- i) Using our alternative measure of ANS acuity, we found that ANS was a concurrent predictor of standardized math scores for Years 0, 2, and 3, but not for Year 1, and that ANS was a concurrent predictor of PC1 across all years.
- ii) Our alternative measure of ANS acuity (hardest ratio tested) predicted PC1 in Years 0 and 1.

iii) A deeper critique, based on the evidence that individuals' scores on different ANS tasks often fail to correlate with one another, is that the non-numerical properties of ANS stimuli (including the controls used, whether trials get increasingly more difficult over time, and the visual properties of the stimuli; Clayton et al., 2015; Inglis & Gilmore, 2014; Smets et al., 2014) may overwhelm the signal of the ANS in ANS assessments. Consistent with this view, even within our dataset, our two measures of ANS acuity differed slightly from one another in predicting math performance.

Funding

Thanks to George Alvarez, for stimulus design and funding support. 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.

Competing Interests

The authors have declared that no competing interests exist.

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 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.

References

- Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences, 21*, 133-137. doi:10.1016/j.lindif.2010.09.013
- Anobile, G., Stievano, P., & Burr, D. C. (2013). Visual sustained attention and numerosity sensitivity correlate with math achievement in children. *Journal of Experimental Child Psychology, 116*, 380-391. doi:10.1016/j.jecp.2013.06.006
- Barner, D., Alvarez, G., Sullivan, J., Brooks, N., Srinivasan, M., & Frank, M. C. (2016). Learning mathematics in a visuospatial format: A randomized, controlled trial of mental abacus instruction. *Child Development*. Advance online publication. doi:10.1111/cdev.12515
- Barth, H. C., & Paladino, A. M. (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. 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. doi:10.1016/j.jecp.2012.09.015
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*, 189-201. doi:10.1037/0012-1649.41.6.189
- Booth, J. L., & Siegler, R. S. (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. M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE, 7*, Article e33832. doi:10.1371/journal.pone.0033832
- 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. doi:10.1016/j.actpsy.2014.01.016

- Chu, F. W., van Marle, K., & Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology*, 132, 205-212. doi:10.1016/j.jecp.2015.01.006
- Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology*, 46, 1176-1191. doi:10.1037/a0019672
- Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica*, 161, 177-184. doi:10.1016/j.actpsy.2015.09.007
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY, USA: Oxford University Press.
- Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*. Retrieved from <http://www.unicog.org/docs/DocumentationDotsGeneration.doc>
- 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. doi:10.1016/j.tine.2013.06.001
- Desoete, A., Ceulemans, A., De Weerd, F., & Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *British Journal of Educational Psychology*, 82, 64-81. doi:10.1348/2044-8279.002002
- DeStefano, D., & LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *The European Journal of Cognitive Psychology*, 16, 353-386. doi:10.1080/09541440244000328
- DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience*, 6, Article 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 calculations in Chinese children: An fMRI study. *Biomed Research International*, 2013, Article 694075. doi:10.1155/2013/694075
- 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. doi:10.1016/j.jecp.2014.01.013
- Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, 141, 134-149. doi:10.1037/a0024427
- Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science*, 16, 136-148. doi:10.1111/desc.12013
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, 47, 1539-1552. doi:10.1037/a0025510
- Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., . . . Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE*, 8, Article e67374. doi:10.1371/journal.pone.0067374

- Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*, 394-406. doi:10.1016/j.cognition.2010.02.002
- Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. *Psychological Science*, *25*, 789-798. doi:10.1177/0956797613516471
- Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (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/a0027433
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 11116-11120. doi:10.1073/pnas.1200196109
- Halberda, J., Mazocco, M. 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*, 47-55. doi:10.1016/0010-0277(77)90016-6
- Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*, 61-83. doi:10.1017/S0140525X0999152X
- Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*, 17-29. doi:10.1016/j.jecp.2008.04.001
- Hornung, C., Schiltz, 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*, Article 272. doi:10.3389/fpsyg.2014.00272
- Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147-155. doi:10.1016/j.actpsy.2013.11.009
- Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic number skills. *Learning and Instruction*, *25*, 95-103. doi:10.1016/j.learninstruc.2012.12.001
- 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. doi:10.1016/j.cognition.2006.10.005
- Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*, 1292-1300. doi:10.1111/j.1467-7687.2011.01080.x
- Libertus, M. E., 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. E., Odic, D., Feigenson, L., & Halberda, J. (2015). *Verbal number estimation predicts math ability and mediates the relation between numerical approximation and math ability*. Manuscript under review, Society for Research on Child Development, Philadelphia, PA, USA.

- Libertus, M. E., 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). The association between higher education and approximate number system acuity. *Frontiers in Psychology*, *5*, Article 462. doi:10.3389/fpsyg.2014.00462
- Link, T., Nuerk, H.-C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *Quarterly Journal of Experimental Psychology*, *67*, 1597-1613. doi:10.1080/17470218.2014.892517
- 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 of the United States of America*, *109*, 18737-18742. doi:10.1073/pnas.1207212109
- Lyons, I. M., & Beilock, S. L. (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. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, *17*, 714-726. doi:10.1111/desc.12152
- Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*, Article e23749. doi:10.1371/journal.pone.0023749
- Mazzocco, M. M. 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
- Moore, A. M., & Ashcraft, M. H. (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
- Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representation of number. *Journal of Experimental Child Psychology*, *103*, 490-502. doi:10.1016/j.jecp.2009.02.003
- 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. *PLoS ONE*, *8*, Article e67918. doi:10.1371/journal.pone.0067918
- 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. doi:10.1016/j.tine.2013.01.001
- 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. doi:10.1016/j.learninstruc.2012.08.004
- Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, *24*, 2013-2019. doi:10.1177/0956797613482944
- Park, J., & Brannon, E. M. (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

- Passolunghi, M. C., Cargnelutti, E., & Pastore, M. (2014). The contribution of general cognitive abilities and approximate number system to early mathematics. *British Journal of Educational Psychology*, *84*, 631-649. doi:10.1111/bjep.12054
- Phelps, G., & Crabtree, S. (2013). *Worldwide, median household income about \$10,000*. Retrieved from Gallup website: <http://www.gallup.com/poll/166211/worldwide-median-household-income-000.aspx>
- 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. 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. doi:10.1126/science.1102085
- Price, G. R., 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., 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
- Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*, 344-357. doi:10.1111/j.2044-835X.2011.02048.x
- Sasanguie, D., Göbel, S. M., 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. doi:10.1016/j.jecp.2012.10.012
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701-703. doi:10.1126/science.171.3972.701
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428-444. doi:10.1111/j.1467-8624.2004.00684.x
- Smets, K., Gebuis, R., Defever, E., & Reynvoet, B. (2014). Concurrent validity of approximate number sense tasks in adults and children. *Acta Psychologica*, *150*, 120-128. doi:10.1016/j.actpsy.2014.05.001
- 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. doi:10.1073/pnas.1302751110
- Stigler, J. W., Chalip, L., & Miller, K. F. (1986). Consequences of skill: The case of abacus training in Taiwan. *American Journal of Education*, *94*(4), 447-479. doi:10.1086/443862
- Sullivan, J., & Barner, D. (2014a). Inference and association in children's early numerical estimation. *Child Development*, *85*(4), 1740-1755. doi:10.1111/cdev.12211
- 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

- Thompson, J. M., Nuerk, H.-C., 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](https://doi.org/10.1016/j.actpsy.2013.05.009)
- Tibber, M. S., Manasseh, G. S. L., Clarke, R. C., Gagin, G., Swanbeck, S. N., Butterworth, B., . . . Dakin, S. C. (2013). Sensitivity to numerosity is not a unique visuospatial psychophysical predictor of mathematical ability. *Vision Research*, *89*, 1-9. doi:[10.1016/j.visres.2013.06.006](https://doi.org/10.1016/j.visres.2013.06.006)
- van Marle, 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. doi:[10.1111/desc.12143](https://doi.org/10.1111/desc.12143)
- Wei, W., Yuan, H., Chen, C., & Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *British Journal of Educational Psychology*, *82*, 157-181. doi:[10.1111/j.2044-8279.2011.02049.x](https://doi.org/10.1111/j.2044-8279.2011.02049.x)
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*, 130-137. doi:[10.1111/1467-9280.00120](https://doi.org/10.1111/1467-9280.00120)
- 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](https://doi.org/10.1016/S0010-0277(99)00066-9)