



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Non-symbolic approximate arithmetic training improves math performance in preschoolers



Joonkoo Park^{a,b,*}, Vanessa Bermudez^c, Rachel C. Roberts^c,
Elizabeth M. Brannon^d

^aDepartment of Psychological and Brain Sciences, University of Massachusetts Amherst, Amherst, MA 01003, USA

^bCommonwealth Honors College, University of Massachusetts Amherst, Amherst, MA 01003, USA

^cCenter for Cognitive Neuroscience, Duke University, Durham, NC 27708, USA

^dDepartment of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:

Received 16 October 2015

Revised 20 July 2016

Available online 2 September 2016

Keywords:

Preschool math

Approximate number system

Cognitive training

Vocabulary

Short-term memory

Executive function

ABSTRACT

Math proficiency at early school age is an important predictor of later academic achievement. Thus, an important goal for society should be to improve math readiness in preschool-age children, especially in low-income children who typically arrive in kindergarten with less mathematical competency than their higher income peers. The majority of existing research-based math intervention programs target symbolic verbal number concepts in young children. However, very little attention has been paid to the preverbal intuitive ability to approximately represent numerical quantity, which is hypothesized to be an important foundation for full-fledged mathematical thinking. Here, we tested the hypothesis that repeated engagement of non-symbolic approximate addition and subtraction of large arrays of items results in improved math skills in very young children, an idea that stems from our previous studies in adults. In the current study, 3- to 5-year-olds showed selective improvements in math skills after multiple days of playing a tablet-based non-symbolic approximate arithmetic game compared with children who played a memory game. These findings, collectively with our previous reports, suggest that mental manipulation of approximate numerosities

* Corresponding author at: Department of Psychological and Brain Sciences, University of Massachusetts Amherst, Amherst, MA 01003, USA.

E-mail address: joonkoo@umass.edu (J. Park).

provides an important tool for improving math readiness, even in preschoolers who have yet to master the meaning of number words.

© 2016 Elsevier Inc. All rights reserved.

Introduction

Mathematical competence during early school years is a powerful predictor of later academic success not only in math but also in other domains such as reading (Aubrey, Godfrey, & Dahl, 2006; Duncan et al., 2007; Geary, Hoard, Nugent, & Bailey, 2013; Jordan, Glutting, & Ramineni, 2010; Jordan, Kaplan, Locuniak, & Ramineni, 2007; Jordan, Kaplan, Ramineni, & Locuniak, 2009; Martin, Cirino, Sharp, & Barnes, 2014). Unfortunately, many children, especially from low-income families, start kindergarten with minimal competencies in math (Jordan, Kaplan, Nabors Oláh, & Locuniak, 2006; Jordan et al., 2007; Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006; Starkey, Klein, & Wakeley, 2004). Thus, enhancing mathematical competence during preschool years is an important societal challenge.

The traditional approach to enhancing mathematical achievement in young children is to target skills foundational to mathematics, including domain-specific abilities such as counting skills and domain-general abilities such as working memory (Baroody, Eiland, & Thompson, 2009; Bryant et al., 2011; Chard et al., 2008; Clements & Sarama, 2007; Dyson, Jordan, & Glutting, 2013; Fuchs et al., 2005; Ramani & Siegler, 2008; Siegler & Ramani, 2009; Starkey et al., 2004). Many of these programs are curriculum based and focus on enhancing specific foundational mathematical knowledge in everyday activities to improve children's mathematical competence. In particular, nearly all of these programs focus on symbolic number concepts such as counting and number facts as well as arithmetic operations with numerical symbols. These intervention programs have started to show great successes in increasing children's mathematical knowledge at young ages (Clements & Sarama, 2011). Nonetheless, there are some limitations to the approaches that currently dominate the landscape. First, many of these programs require a large investment of teacher training and must be implemented over a full academic year. Relatedly, because the teachers are given a great deal of additional training to implement the intervention curricula, it is often difficult to find a fair control group. Another limitation from the standpoint of cognitive science is that most interventions are very heterogeneous, often consisting of multiple different tasks or skills. Although such large-scale interventions may be a very effective strategy for improving mathematical ability, it is difficult to pinpoint which aspects of the curriculum are most effective.

Here we take an alternative approach and ask whether training aimed at children's preverbal number sense could be an effective way to improve math readiness. Children come into the world with a preverbal number sense that allows an approximate sense of quantity (Dehaene, 1999; Gallistel & Gelman, 1992). This system has been termed the approximate number system (ANS) and has been shown to be present in newborn infants (Izard, Sann, Spelke, & Streri, 2009; Xu & Spelke, 2000), to operate in adults from cultures without symbolic counting systems (Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2004), and to be present in a wide range of nonhuman animals (Agrillo, Cohen Kadosh, & Dowker, 2014; Beran, Perdue, & Evans, 2015). Furthermore, the system increases in precision with age and is found across the lifespan (Halberda & Feigenson, 2008; Halberda, Ly, Wilmer, Naiman, & Germine, 2012).

Despite the profound differences between a symbolic representation of number, which allows the representation of number with extreme precision, and approximate non-symbolic representations, which are noisy and follow Weber's law, recent studies have uncovered a correlation between individual ANS acuity and symbolic math ability (e.g., Gilmore, McCarthy, & Spelke, 2010; Halberda, Mazzocco, & Feigenson, 2008; Libertus, Feigenson, & Halberda, 2011; Mazzocco, Feigenson, & Halberda, 2011; but see Fuhs & McNeil, 2013; Gobel, Watson, Lervag, & Hulme, 2014; Inglis,

Attridge, Batchelor, & Gilmore, 2011; Kolkman, Kroesbergen, & Leseman, 2013; Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013; Sasanguie, Gobel, Moll, Smets, & Reynvoet, 2013; Tibber et al., 2013; Wei, Yuan, Chen, & Zhou, 2012). We use the term *symbolic math* to refer to any *learned* mathematical concepts and knowledge, including identifying symbols, number words, using the count list, and solving arithmetic problems based on number words or numerals. This correlation between the ANS and symbolic math indicates that approximate numerical abilities may be foundational for learned math skills and, therefore, suggests an alternative approach to improving math readiness. However, given the correlational nature of the findings, alternative possibilities are plentiful (e.g., see Gilmore et al., 2013; Lyons & Beilock, 2011; Piazza, Pica, Izard, Spelke, & Dehaene, 2013). To investigate the causal nature of the relationship, our previous research asked whether repeated training on an approximate arithmetic task, which requires addition and subtraction of non-symbolic numerical quantities, would selectively improve mental arithmetic skills in college students (Park & Brannon, 2013, 2014). In these experiments, adult participants first received a set of pretests, which included a symbolic arithmetic test, among others. Participants were then randomly assigned into one of multiple cognitive training groups. In each study, the experimental training group spent 6 to 10 daily sessions engaged in approximate non-symbolic arithmetic, whereas control groups underwent other types of cognitive training such as visual short-term memory and numerosity comparison training. We found that participants who were engaged in the non-symbolic approximate arithmetic training—and only those participants—showed improvement in symbolic arithmetic skills. Given that numerosity comparison and short-term memory training did not yield significant benefit in symbolic arithmetic performance, we concluded that mental manipulation of nonverbal numerical quantity was the critical factor driving improvements in math skills. Consistent with our findings, another research group found that first graders briefly exposed to similar preverbal number sense tasks performed better than peers in a subsequent symbolic arithmetic test (Hyde, Khanum, & Spelke, 2014). Collectively, these recent findings suggest that engaging the preverbal number sense via approximate arithmetic may be an effective way to improve mathematical ability throughout the lifespan.

Here, we tested this possibility by asking whether approximate arithmetic training benefits preschoolers' math abilities. Local preschoolers, largely from low-income families, first received a battery of pretests that assessed their math, vocabulary, verbal short-term memory, and executive function. Children were then pseudorandomly assigned to one of two groups; the first group played a tablet-based game that required adding and subtracting arrays of icons, and the second group played a tablet-based game that required remembering the location of pairs of pictures. Then, we assessed the effectiveness of the two cognitive training programs by comparing pretest and posttest scores for each of the test measures. From this study design, we aimed to evaluate how approximate arithmetic—presumably the process of mental manipulation of numerical quantity—could benefit very young children at the verge of the introduction to formal math education and often with minimal symbolic math knowledge. This study was meant as an important first step to test the effectiveness of approximate arithmetic training in preschool children with no formal math education and as a springboard for a larger scale math intervention study.

Method

Participants

Participants were recruited from local preschools that were largely composed of children funded by a state-funded preschool program. This program provides assistance to children from families whose gross income is at or below 75% of the state median income level. We intentionally focused on children of low-income families because they represent the greatest need for improving early mathematical competence. A total of 109 preschoolers across four schools initially enrolled in the study.

Children must have completed pre- and posttest sessions and all 10 training sessions to be included in the final analysis. Of the 109 total children, 6 did not meet these requirements and, thus, were excluded. Of these 6 children, 4 were from the math group and 2 were from the memory group (see "Procedure" section below). One of the children finished the training phase but was unable to perform the posttests because she moved to a different school. The remaining 5 children completed

only 1 to 6 training sessions because they moved to a different school ($n = 2$), because of extended absence due to illness ($n = 1$) or family travel ($n = 1$), or because of unwillingness to play the game ($n = 1$). The final sample consisted of 103 preschoolers ranging in age from 3.94 to 5.56 years ($M = 4.87$ years, $SD = 0.40$; 53 boys). Most children were from low-income families who qualified for state-funded preschool, whereas 24 of the children were from middle-income families enrolled in classrooms in the same preschools. Tests were administered in Spanish by a native Spanish speaker for 16 of the bilingual children who appeared to be more comfortable speaking in Spanish.

Parents gave written consent for their children to participate in the study. Children were compensated with stickers and small toys throughout the study and with a math-themed storybook and a building toy at the completion of the study. In addition, each participating classroom received an educational toy of the teacher's choice. All procedures were approved by Duke University's institutional review board.

Procedure

The study consisted of three phases: pretest, training, and posttest. Pre- and posttests were conducted individually for each child in a small room at each preschool location. Pretests were administered over two sessions of approximately 25 to 35 min each. In Session 1, children completed the Test of Early Mathematical Achievement–Third Edition (TEMA-3), a standardized test for assessing mathematical achievement. In Session 2, children completed a receptive vocabulary assessment (Peabody Picture Vocabulary Test–Fourth Edition, PPVT-4), a short-term memory task, and two executive function tasks (Stroop interference and card sorting) in that order. Posttests were administered the same way except that children were assessed on different versions of the same tests. All five assessments had two versions (e.g., Versions A and B). The order of the versions (pretest vs. posttest) was counter-balanced across children.

After completing the pretests, children in a given classroom were pseudorandomly assigned to either the approximate arithmetic group or the picture memory group with a constraint to minimize differences in age, sex, and TEMA and PPVT pretest scores between the two groups. To be specific, all children in the given classroom were randomly assigned to two groups, and this random assignment was done 10,000 times. In each of the 10,000 repetitions, a t -test was performed on age, sex, and TEMA and PPVT pretest scores, and the average p -values for age and sex were recorded only when both TEMA and PPVT p -values were larger than .95. Then, the random grouping that yielded the largest average p -value (of age and sex) was selected as the final group assignment. The randomization process was done by the first author, who never directly interacted with the children.

The training phase, which consisted of 10 12-min sessions, was conducted in small groups of 2 to 4 children. The 10 daily sessions were typically spread out over 2 to 3 weeks. Children wore headphones to prevent auditory interference while listening to verbal instructions. During the training phase, the approximate arithmetic group completed an adaptive tablet-based game (using an iPad; Apple) requiring children to approximately add and subtract large arrays of icons. The picture memory group completed an adaptive tablet-based card-matching game where children flipped and matched pairs of cards. The training phase was followed by posttests. The logistics of the project prevented us from relying solely on experimenters who were blind to children's experimental group assignments. However, 17 of the children were blind-tested. It took approximately 1 month to complete the study in each classroom.

Approximate arithmetic training

The approximate arithmetic group training game (Fig. 1A) was modeled after the approximate arithmetic training task used in previous studies with adults (Park & Brannon, 2013, 2014). On each trial, children viewed an animation of addition or subtraction with two arrays containing discrete objects (e.g., paintbrushes, palm trees). Children were asked to help a fictitious character, Max, choose the right answer.

Trials followed one of four different structures: addition compare, addition match, subtraction compare, or subtraction match. On all trials, children were first asked to “touch the red balloon,” at

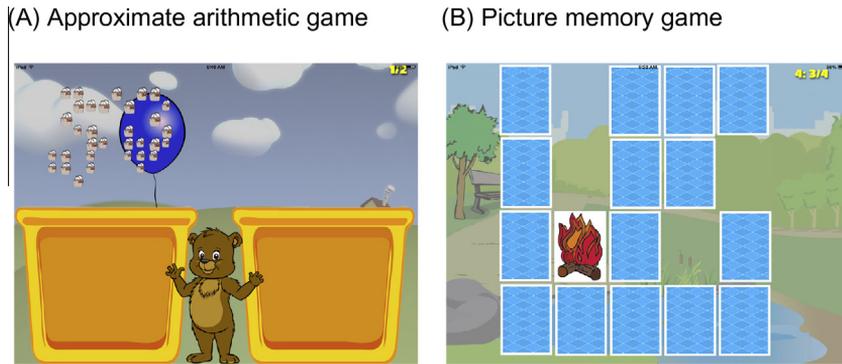


Fig. 1. Screen shots of the two training games.

which point the balloon popped and an array of items dropped into an opaque bucket with a verbal remark, “Look, some [items] are going into the box.” On the addition trials, the blue balloon then popped spontaneously and a second array of homogeneous items dropped into the opaque bucket with a verbal remark, “More [items] are going into the box.” On the subtraction trials, after the contents of the red balloon dropped into the opaque bucket, a bird spontaneously popped out of the blue balloon and took a subset of the items away and off the screen with a verbal remark, “Some [items] are taken away.” In both cases, children were instructed to imagine the total number of items in the opaque bucket. On addition compare and subtraction compare trials, a second visible array was shown on the right side of the opaque bucket, and children were verbally instructed to “touch the one that has more” after the addition or subtraction event was complete. This second array contained a number of items that differed by a specific ratio (or a difference in a \log_2 scale) to the correct sum or difference (either larger or smaller than the correct answer). On addition match and subtraction match trials, two additional arrays were shown on the right side of the screen. Here, children were verbally asked to “touch the one that shows Max how many he has.” One of these two arrays contained the correct sum or difference that matched the contents of the opaque bucket, and the other contained a number that differed by a specific ratio (either larger or smaller than the correct answer).

Stimuli remained on the screen until a response was made. When children made a correct response, they heard one of the following responses: “nice job,” “you are correct,” or “yes, that’s right!” randomly chosen. After an incorrect response, they heard “nope,” “try again,” or “not quite” randomly chosen. The four trial types were presented in blocks of 10 trials, and the order of the blocks was fixed from addition compare, subtraction compare, addition match, to subtraction match. This order continued throughout the training. For example, if a child played 5 blocks in one session ending with addition compare, then the child’s next session started from subtraction compare. Children on average played a total of 40 blocks (roughly 4 blocks per session) throughout the experiment.

The number of items in each array ranged from 4 to 80. On approximately 87% of the trials, the number of objects in the two arrays representing the operands ranged from 4 to 16 items in the addition conditions and from 12 to 40 in the subtraction conditions. This design yielded correct answers ranging from 12 to 20 in both the addition and subtraction conditions. This narrow number range for the correct answers was necessary because we started the game with the \log_2 difference of 2 (4:1 ratio), and the maximum number of objects that could fit into the container was limited to 80 due to the screen size of an iPad. Nevertheless, because all of the correct answers were within this range, it would be technically possible for participants to learn the structure of the problems and answer based just on the final array presented without actually adding or subtracting the first two arrays, especially when the \log_2 difference level is high. To minimize the use of such a strategy, we incorporated another problem structure on the remaining 13% of the trials. On these trials, the addition and subtraction problems were designed so that the correct answers ranged widely from 4 to 56 items, although this problem structure limited the alternative choices (e.g., if the correct answer was 4

and the \log_2 difference was 2, then the alternative option was always 16 and was never 1; if the correct answer was 56 and the \log_2 difference was 2, then the alternative option was always 14 and was never 224).

Difficulty was manipulated separately for each trial type by varying the numerical difference between the correct and incorrect answers on a log scale, beginning with a log difference of 2 for the easiest level (e.g., 12 vs. 48 items). Following a typical psychophysical adaptive staircase procedure, the \log_2 difference level increased by one of the values randomly chosen from [0.08, 0.09, 0.10, 0.11, 0.12] when the average accuracy of a block of 10 trials was equal to or less than 65%, and the level decreased by one of the values randomly chosen from [0.13, 0.14, 0.15, 0.16, 0.17] when the average accuracy was equal to or greater than 80% (Park & Brannon, 2013, 2014). After every 2 blocks (i.e., 20 trials), children were shown a short classic children's cartoon (45–60 s) that was designed to maintain interest and motivation.

Picture memory training

Children in the picture memory group played an iPad game that began with a virtual set of face-down cards (Fig. 1B). Children were verbally instructed to “flip the cards over to make a pair.” Touching a card resulted in it flipping over to reveal an icon that remained visible until a second card was touched. If the icons matched, then both cards disappeared from the board with a sweeping bell tone. In contrast, if the two icons did not match, then both cards returned to their original face-down position without a sound. The game comprised seven different difficulty levels based on the number of cards presented: 2×2 , 3×2 , 4×3 , 5×4 , 6×5 , 7×6 , and 8×7 cards. The difficulty level for each child was set each day such that the child would play easier levels before progressing to more difficult levels on that day. For the majority of children Levels 2×2 and 3×2 were very easy, and for all children Levels 7×6 and 8×7 were very difficult; therefore, nearly all children played trials set at Levels 4×3 , 5×4 , and 6×5 daily. Note that picture memory training was not adaptive in the sense of the approximate arithmetic game; however, the progression of difficulty levels within each day allowed children to be challenged and engaged without becoming too frustrated and discouraged. Children were shown the same 45- to 60-s video cartoon clip as in the approximate arithmetic group after roughly every 4 min of game play to maintain children's interest in completing the session.

Math assessment

We measured symbolic mathematical ability with the third edition of the Test of Early Mathematical Achievement (Ginsburg & Baroody, 2003). This test, which is designed for children age 3 years 0 months to 8 years 11 months, assesses formal (e.g., numeral literacy) and informal (e.g., relative magnitude) mathematics abilities. Examples of questions primarily targeted for ages 4 and 5 years include asking children to verbally count from 1 to 10, to hold up four fingers, to state which number (word) comes after a single-digit number, to choose the numerically larger of the two Arabic numbers, and to solve simple single-digit addition and subtraction word problems. For Spanish-speaking students, directions were given in Spanish and answers in either English or Spanish were acceptable. Raw scores were obtained from all children, and scores were converted to a normalized scale based on each child's age at the time of enrollment in this study, with a standard score of 100. It should be noted that although TEMA-3 includes a very small number of nonverbal math problems (e.g., asking a child to reproduce a set of items to match based on number), the majority of the problems are symbolic and verbal in nature and none of the problems requires nonverbal estimation or addition and subtraction of large values.

Reliability measures for the TEMA-3 are known to be relatively high, ranging from .82 to .97 across different types of reliability coefficients, according to the examiner's manual. In addition to these published measures, our pretest to posttest design allowed a unique opportunity to assess the correlation between pretest and posttest TEMA measures across children within each training group, which serves as a proxy for a test–retest reliability measure. The correlation coefficients were $r = .85$ and $r = .94$ in the approximate arithmetic group and the picture memory group, respectively (see Fig. 4 below in Results), suggesting reasonably high reliability of the TEMA scores.

Vocabulary assessment

Children's receptive vocabulary was assessed using the fourth edition of the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). The test is designed for children as young as 2.5 years. On each trial, the experimenter presented four pictures, spoke one word, and asked the child to select which of the pictures that word described. For Spanish-speaking students, the word was spoken in both English and Spanish. Raw scores were obtained from all children and converted to a normalized scale with a standard score of 100 based on each child's age at the time of enrollment in this study.

Published reliability measures of the PPVT are very high, ranging from .88 to .95 across different types of reliability coefficients, according to the examiner's manual. As in the case of the TEMA, we assessed the correlation between pretest and posttest PPVT measures across children within each training group in order to get a rough estimate of the reliability of the measures in our sample. The correlation coefficients were $r = .80$ and $r = .83$ for the approximate arithmetic group and the picture memory group, respectively, also suggesting reasonable reliability of the PPVT scores (see Fig. 4 below).

Short-term memory assessment

A letter span task was used to assess children's verbal short-term memory (STM). On each trial, the experimenter asked the child to repeat a string of letters in the correct order. Children were tested with up to six different string lengths starting with two-item sequences and progressively increasing to seven-item sequences. Children were given a total of 5 trials at each string length, and the testing was terminated when participants erred on 3 or more of the 5 trials. Only 13 letters of the alphabet were used, and different letters for Spanish and English speakers were used to ensure that all letters were monosyllabic. Similar sounding letters such as M and N, D and B, and P and T in English were excluded. Total number of correctly repeated strings was used as the final dependent measure. The correlation between pretest and posttest STM measures across children within each training group were $r = .83$ and $r = .82$ in the approximate arithmetic group and the picture memory group, respectively, indicating reasonable reliability (see Fig. 4 below).

Stroop interference test

A measure of children's executive function (EF) was constructed from two tasks: a Stroop interference test and a card sorting test. First, a variant of the Stroop-like "day–night" task (Gerstadt, Hong, & Diamond, 1994) was used to probe inhibitory control, a component of executive function. "Day–night" was replaced with "red–blue" because pilot testing on an independent group of preschoolers revealed that children often responded inconsistently to the original images (e.g., by saying "moon" in place of "night"). This test consisted of 16 congruent and 16 incongruent trials. For children who did not know the color names (6 children), we conducted the same Stroop interference task with "cat and dog" instead of "red and blue."

On congruent trials, the experimenter presented a red or blue square (alternatively a cat or dog) and the child was asked to label the card as red or blue (alternatively cat or dog). Children were required to answer correctly on four practice trials before starting the 16 congruent trials successively. Children were then told that they would play a "silly version" of the game, the incongruent trials, in which they would name the opposite of each card—for example, say "red" for a blue card. Children practiced this silly version until they correctly answered four practice trials and were then given 16 incongruent trials successively.

An audio recording was made for later analysis of accuracy and reaction time. The experimenter said the word "go" every time a card was turned over in order to mark the moment each card became visible to the child. Reaction time for each trial was measured from the beginning of the experimenter's word "go" to the beginning of the child's response. Self-correction was ignored, and only the first response was evaluated (e.g., "red ... no, blue" to a red card in the opposite trial would be marked as incorrect). Two card orders were created so that children received different versions for pretests and posttests. The order of the versions administered was counterbalanced across participants.

Performance was assessed by a composite score in which the number of correct responses on incongruent trials was divided by the number of correct responses on congruent trials. A composite score of 1 represents equal performance between the congruent and incongruent trials. Composite scores of less than 1 represent a decrement in performance in the incongruent trials. A second analysis incorporated reaction time. The number of correctly responded incongruent trials was divided by the median reaction time of all the incongruent trials, and the same measure was computed from the congruent trials. The ratio between the two measures served as the composite score. The overall results from this analysis were qualitatively identical to the original analysis (i.e., the gain score difference between the two experimental groups did not show a reliable difference, $t(97) = -0.373$, $p = .710$); therefore, we report the results from the original analysis.

Unlike the other pre- and posttest measures described so far, this measure yielded poor reliability. To be specific, the correlation coefficients between pretest and posttest measures were $r = -.08$ and $r = .56$ in the approximate arithmetic group and the picture memory group, respectively (see Fig. 4 below).

Card sorting test

A variant of the Standard Dimensional Change Card Sort task (Zelazo, Frye, & Rapus, 1996) was used to probe executive function. Two versions (Version A with bird/fish and Version B with plane/ship) were created to avoid administering the same task for both pre- and posttests, and the order of the versions was counterbalanced across participants. Each version consisted of a set of 10 cards that could be sorted according to two rules: color (Rule 1: white/black) and shape (Rule 2: bird/fish or plane/ship depending on the version). In Version A, there were 5 white cards and 5 black cards that could be divided into 6 bird cards and 4 fish cards. In Version B, there were 6 white cards and 4 black cards that could be divided into 5 planes and 5 ships.

Two boxes were labeled with a target card (e.g., one box with a white bird and one box with a black fish) that needed to be matched according to the rule of each trial (e.g., color for Rule 1). Children first watched the experimenter sort 3 cards into boxes according to Rule 1 (color) and then were asked to sort 10 additional cards as quickly as possible. The experimenter then demonstrated sorting 3 cards according to Rule 2 (shape) and asked children to sort the same 10 cards according to the new rule. Accuracy and time spent to complete each trial were recorded.

A composite score was computed by dividing the number of correctly sorted cards according to Rule 2 by the number of correctly sorted cards according to Rule 1. The gain scores (posttest composite score minus pretest composite score) in this test resulted in a large variation, indicating potential outliers. Accordingly, we removed extreme outliers that were 2.5 times the interquartile range from the first and third quartiles. This procedure resulted in five outliers: one from the approximate arithmetic group and four from the picture memory group. We performed a second analysis taking the total time spent sorting into account (by dividing the number of correctly sorted cards by the total time spent in each rule), and the results did not differ qualitatively from our initial analysis (i.e., the gain score difference between the two experimental groups did not show a reliable difference, $t(99) = -0.214$, $p = .831$).

Similar to the Stroop interference test, the card sorting test also yielded poor reliability. The correlation coefficients between the pretest and posttest measures were $r = .11$ and $r = .39$ for the approximate arithmetic group and the picture memory group, respectively (see Fig. 4 below). A summarized measure of EF, the sum of Stroop interference and card sorting scores, did not improve the reliability by much ($r = .10$ and $r = .51$). Note that questionable reliability in the EF tasks makes it difficult to interpret any changes from pretest to posttest as a function of training.

Results

We first report the performance progression in the two experimental games. Children who played the approximate arithmetic game showed a consistent decrease in the log difference level (Fig. 2A), as indicated by a negative slope of the linear fit for each individual child's log difference level,

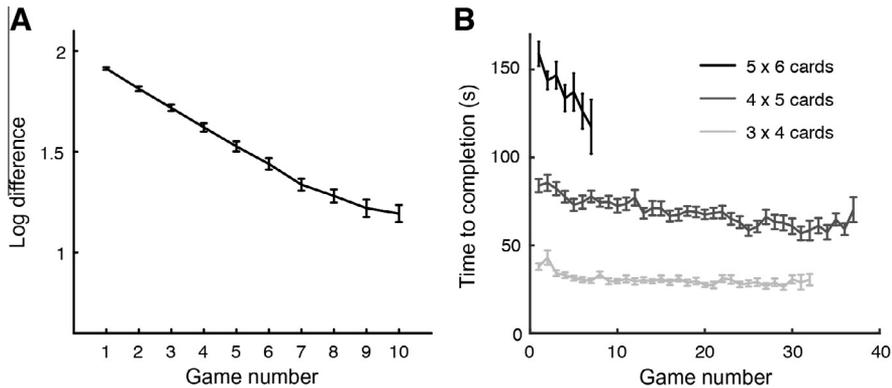


Fig. 2. Progression of the dependent measure over the course of training on the approximate arithmetic group (A) and the picture memory game (B). A game in the approximate arithmetic game represents one set of 40 trials, and the corresponding log difference value indicates the \log_2 difference value at the end of each game. A game in the picture memory game represents one round of an n -by- n card game. Error bars represent the standard error of the mean.

$t(50) = -20.391, p < .001$. This monotonic decrease in the log difference level could reflect the progression of the threshold of the adaptive staircase procedure used in this game rather than children's improved performance in approximate arithmetic (Lindskog & Winman, 2016). To test whether there is direct evidence for performance improvement in this adaptive training design, we adopted a novel analytic technique that quantified the trend in accuracy under comparable difficulty levels across time (Park & Brannon, 2016). In each participant, the accuracy measure from the median log difference level (from 45th to 55th percentiles) across all blocks was regressed on time (session). The slope of this regression model was nonsignificantly positive (slope = 0.0183), $t(11) = 0.673, p = .515$, suggesting that there is no direct evidence for improvement in approximate arithmetic performance over the training sessions. However, as elaborated in Park and Brannon (2016), the treatment effect of cognitive training does not hinge on evidence of observable performance improvement in the training task because the effect of treatment is through improvement, facilitation, strengthening, or growth of the unobservable cognitive element that drives the transfer of skills between the training and the target tasks. Children who received the picture memory game showed significant reduction in time to completion (Fig. 2B), particularly in more difficult conditions, perhaps due to a floor effect in the easiest condition (Level 3: linear effect, $t(51) = -1.593, p = .117$; Level 4: linear effect, $t(49) = -2.719, p = .009$; Level 5: linear effect, $t(32) = -3.905, p < .001$). Overall, results from both of these training tasks suggest that young children were capable of performing the tasks consistently over the course of the experiment.

The critical analyses central to our research question were the comparisons of the pretest to posttest scores on the TEMA, PPVT, STM, and EF assessments in the two experimental groups. In Table 1, we report the pre- and posttest scores along with participant demographics in each group. Our pseudo-random group assignment ensured that there was little difference in the TEMA pretest scores, $t(101) = 0.311, p = .756$, and the PPVT pretest scores, $t(101) = 0.508, p = .613$, between the two groups. Pretest scores in STM, $t(101) = 1.289, p = .200$, Stroop interference, $t(97) = -0.647, p = .519$, and card sorting, $t(99) = 0.184, p = .854$, also showed few group differences.

The effects of approximate arithmetic and picture memory training on pretest to posttest changes were evaluated first using the gain score approach and then using the analysis of covariance (ANCOVA)/regression-based approach.

Standardized gain scores were computed by taking the difference between each of the post- and pretest measures and dividing this value by the standard deviation of the pretest measures. The gain scores for the two training groups were compared using two-sample t -tests (Fig. 3). Among all the pre- and posttest measures, TEMA gain scores were found to be significantly greater in the approximate arithmetic group (gain score of 4.549) compared with the picture memory group (gain score of

Table 1
Mean pre- and posttest scores (and standard deviations) sorted by the two training groups.

	Age (years)	Sex (# boys)	TEMA		PPVT		STM		Stroop interference		Card sorting	
			Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
			(12.1)	(14.1)	(13.6)	(15.4)	(3.2)	(3.3)	(0.23)	(0.24)	(0.36)	(0.28)
Approximate arithmetic group (n = 51)	4.9 (0.4)	26	95.5 (12.1)	100.1 (14.1)	96.6 (13.6)	98.6 (15.4)	11.8 (3.2)	12.1 (3.3)	0.60 (0.23)	0.64 (0.24)	0.95 (0.36)	0.83 (0.28)
Picture memory group (n = 52)	4.9 (0.4)	27	94.6 (17.2)	96.4 (16.3)	95.1 (16.8)	95.2 (15.9)	10.8 (4.3)	12.0 (4.2)	0.63 (0.23)	0.69 (0.23)	0.94 (0.42)	0.83 (0.36)

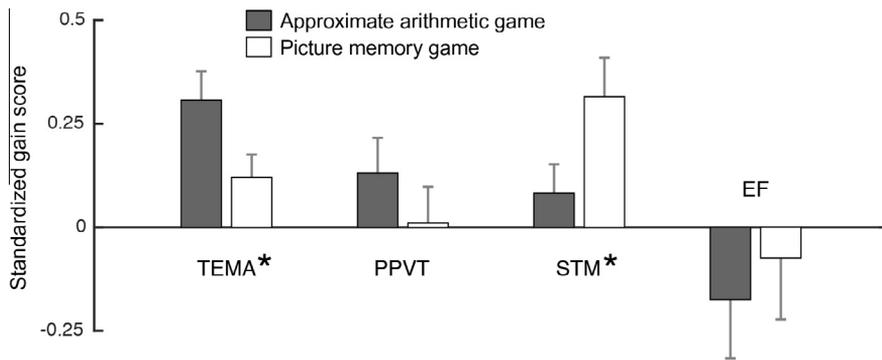


Fig. 3. Mean standardized gain scores of multiple assessments. Error bars indicate the standard error of the mean. * $p < .05$.

1.789) (Cohen’s $d = 0.414$ computed with a pooled standard deviation of 6.670), $t(101) = 2.100$, $p = .038$, and STM gain scores were found to be significantly greater in the picture memory group compared with the approximate arithmetic group, $t(101) = -1.990$, $p = .049$. There were no group differences in the gain scores in the PPVT measure, $t(101) = 0.991$, $p = .324$, or in the EF measure, $t(95) = -0.477$, $p = .635$.

Although the gain score results (Fig. 3) provide a preliminary view of the effects of cognitive training (i.e., treatment) on pretest to posttest change, particularly in TEMA and STM, the gain score approach is not without its weaknesses; for example, it does not adequately take into consideration potential individual differences in the pretest scores and is likely to have large statistical noise, especially if the measures are not reliable. Thus, an ANCOVA/regression-based approach was performed in which posttest scores were predicted by categorical group membership (approximate arithmetic vs. picture memory) with the pretest scores entered as a covariate. The interaction between group and pretest was initially modeled but was then dropped if it did not reach significance at $p < .10$. The interaction was dropped in the model for TEMA, PPVT, and STM, whereas it was retained in the model for EF. Note that the beta estimate associated with group membership represents the point estimate of the treatment effects if divided by 15, which is the assumed population standard deviation of the normed TEMA and PPVT scores.

There was a significant treatment effect in TEMA scores ($\beta = 2.83$, $\eta_p^2 = .045$), $F(1, 100) = 4.73$, $p = .032$ (Fig. 4). This pattern held when only the blind-tested children were analyzed ($\beta = 7.40$, $\eta_p^2 = .283$), $F(1, 14) = 5.54$, $p = .034$. More central to the aim of this study, we performed the same analysis only with children from low-income families. In these children, TEMA scores improved from 93.3 to 98.7 in the approximate arithmetic group and from 93.2 to 95.5 in the picture memory group, which resulted in a significant treatment effect, with a somewhat larger effects size than when all children were included ($\beta = 3.21$, $\eta_p^2 = .056$), $F(1, 76) = 4.49$, $p = .037$. We also performed the same analysis separately on younger and older age groups split by the median age (4.9 years). There was a significant

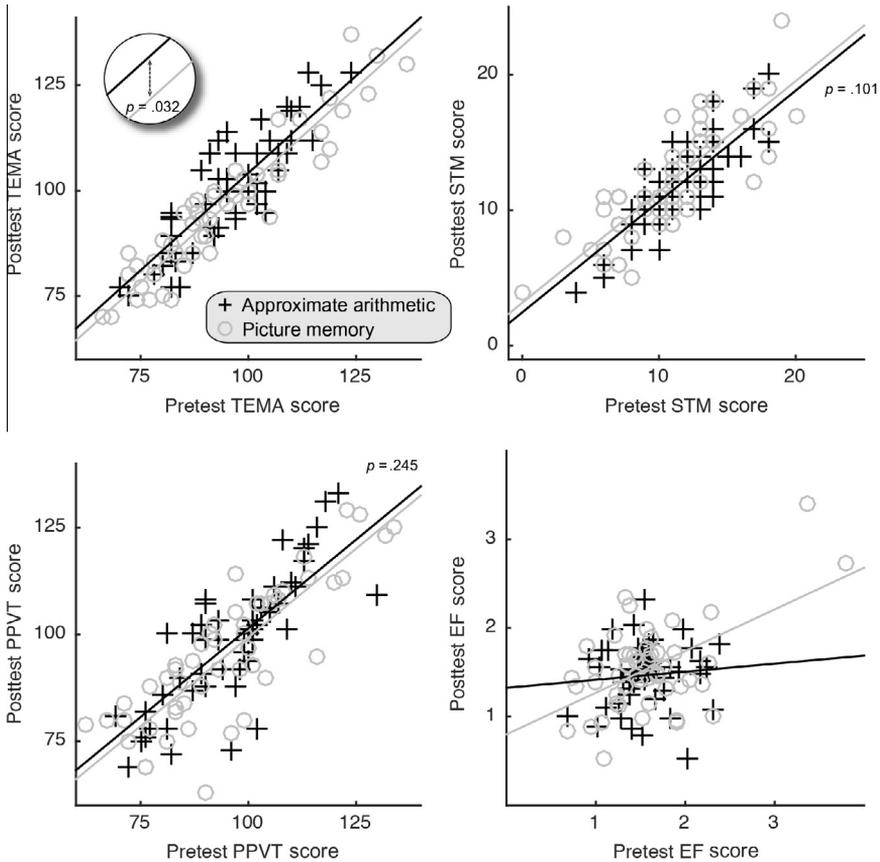


Fig. 4. Results of the ANCOVA/regression-based approach to assess the treatment effects in the TEMA, PPVT, STM, and EF measures.

treatment effect in younger children ($\beta = 4.09$, $\eta_p^2 = .091$), $F(1, 48) = 4.83$, $p = .033$, but not in older children ($\beta = 1.22$, $\eta_p^2 = .010$), $F(1, 48) = 0.49$, $p = .485$. Finally, the treatment effect on TEMA remained significant when all of the pretest measure scores (PPVT, STM, and EF) were included as covariates ($\beta = 2.70$, $t(92) = 2.02$, $p = .046$).

The same ANCOVA/regression-based approach was performed on PPVT and STM scores (Fig. 4). The treatment effect in PPVT was not significant, although the effect size in terms of beta coefficient was moderate ($\beta = 2.11$, $\eta_p^2 = .013$), $F(1, 100) = 1.37$, $p = .245$. The treatment effect in STM was not significant, unlike the gain score results ($\beta = -0.70$, $\eta_p^2 = .027$), $F(1, 100) = 2.73$, $p = .101$, which suggests that significant group differences in the STM gain scores were perhaps due to added noise and/or idiosyncratic individual differences in the pretest measures. There were no treatment effects on EF scores ($\beta = -0.07$, $\eta_p^2 = .007$), $F(1, 94) = 0.66$, $p = .419$, whereas the interaction was significant ($\beta = 0.38$), $t(93) = 2.09$, $p = .040$. Collectively, the ANCOVA/regression-based results (as well as the gain score results) indicate that playing the approximate arithmetic group game compared with playing the picture memory game selectively improved TEMA scores, and no other measures showed a significant improvement.

One unexpected finding was that the approximate arithmetic game appeared to yield a non-significant but small benefit in PPVT performance ($\beta = 2.11$). In the approximate arithmetic game, the objects are repeatedly named on each trial (e.g., “Look, some donkeys are going into the box,” “More donkeys are going into the box”). In contrast, there was no verbal labeling of the items in

the picture memory game. Repeated labeling of object identity in the verbal instructions of the approximate arithmetic game, therefore, may have enhanced children's name labeling skills, which is central to the PPVT. At a more concrete level, we also considered the possibility that items labeled in the approximate arithmetic game were actually tested in the PPVT. For instance, a child who initially does not know what a donkey is may have learned what it looks like and what it is called by playing the approximate arithmetic game, thereby resulting in an improved PPVT score when the child is asked to find a "donkey" at posttest. In fact, there were five words on each of the two PPVT versions that appeared in the approximate arithmetic game. An important question, then, is to what extent these five words contributed to the PPVT scores. We used the following approach to quantify this contribution. First, the contribution of the five words (that appeared in the approximate arithmetic game) in each child's posttest PPVT score was computed by taking the total number of words tested and got correct (0–5)¹ divided by the child's raw PPVT score. Across children, this value (let us call it k) ranged from 0 to 0.1739 with a median of 0.0147. Then, each child's normed posttest PPVT score was adjusted to remove the contribution of those words by multiplying $1 - k$ to the posttest score. Finally, an ANCOVA was used to test the treatment effect of experimental group on this adjusted posttest score with the pretest score as a covariate. The treatment effect was smaller in effect size ($\beta = 2.01$, $\eta_p^2 = .009$), $F(1, 100) = 0.936$, $p = .336$, than the treatment effect of the original analysis without removing the contribution of the five words ($\beta = 2.11$, $\eta_p^2 = .013$), $F(1, 100) = 1.37$, $p = .245$. These results are consistent with the idea that verbal labeling in the approximate arithmetic game may have influenced the posttest PPVT scores more positively than in the picture memory game.

One may wonder whether there exist any ANS-based problems in the TEMA and whether improved TEMA after approximate arithmetic training may have been driven by improvement solely in those ANS-based problems. It should be noted that nearly all of the TEMA problems are symbolic (i.e., language based) in that they involve counting, enumeration, and numerals, and only one problem ("perception of more," which is akin to a non-symbolic numerosity comparison task) is potentially ANS based. Thus, it is unlikely that the main treatment effect in TEMA is driven by that single problem. Nevertheless, we provide empirical support to this argument as follows. Given that the TEMA is a heterogeneous assay of number knowledge, we quantified itemized gain scores for each participant for each item and computed the mean gain score for each problem item (from Problem 1 to Problem 32) across all children in the approximate arithmetic group.² Following the primary finding, these gain scores were significantly different from zero, $t(31) = 5.503$, $p < .001$, $d = 0.949$. The same analysis done in the picture memory group resulted in a significant gain score but with a much smaller effect size, $t(31) = 2.078$, $p = .046$, $d = 0.358$. Indeed, the itemized gain scores were greater in the approximate arithmetic group than in the picture memory group, $t(31) = 2.497$, $p = .018$. Now, this statistic was computed again after excluding the gain score from that one ANS-based problem, and the gain score remained highly significant in the approximate arithmetic group, $t(30) = 5.452$, $p < .001$, $d = 0.955$. The same analysis performed in the picture memory group again resulted in a much smaller effect size, $t(30) = 1.944$, $p = .061$, $d = 0.340$, which was comparatively smaller than the effect of the approximate arithmetic group, $t(30) = 2.580$, $p = .015$.

Discussion

Although humans are uniquely capable of symbolic mathematics, we share with other animals an intuitive ability to estimate, compare, and manipulate large numbers of items without counting that is present during human infancy before symbolic math skills are acquired (Dehaene, 1999). A growing body of research suggests that the preverbal number sense that supports approximate non-symbolic number representations is fundamentally related to full-fledged math abilities, particularly

¹ The starting point of PPVT questions depends on the child's age, and the child is progressively probed until he or she makes a certain number of wrong answers. Thus, if a word is (thought to be) too easy or too difficult for a child, the child does not get tested on those words.

² Itemized gain score of 1 if a participant got the problem incorrect in pretest but correct in posttest, -1 if the participant got the problem correct in pretest but incorrect in posttest, and 0 otherwise. Note that the TEMA assesses math skills through progressive probes until the participant makes a certain number of wrong answers. In our study, the 75% percentile of the highest problem number solved in all children was 32.

in children (Fazio, Bailey, Thompson, & Siegler, 2014; Gilmore et al., 2010; Halberda et al., 2008; Libertus et al., 2011; Mazzocco et al., 2011). However, the nature of this relationship and its developmental trajectory are being queried by many research groups and proposals have been made that attribute the correlation between the ANS and math to tertiary mediating factors. Thus, research approaches that directly test the causal relationship between them are sorely needed. To that end, our previous studies demonstrated that repeated training on approximate arithmetic, one form of intuitive number sense, improves mental arithmetic skills in adults (Park & Brannon, 2013, 2014). Importantly, similar training tasks designed to enhance approximate comparison or matching, visuospatial short-term memory, or a numeral order judgment did not show such an effect (Park & Brannon, 2014). From these previous findings, we proposed that numerical quantity *manipulation* in the mental workspace may be a common cognitive denominator to both non-symbolic approximate arithmetic and symbolic mental arithmetic, and that facilitation of this common denominator via approximate arithmetic training may drive effective cognitive processing for solving mental arithmetic problems (Park & Brannon, 2014).

The goal of the current study was to test the effectiveness of approximate arithmetic training in young children with minimal knowledge in math. Testing such an effect in young children is critical for both theoretical and practical reasons. Theoretically, it allows more in-depth assessment of the causal link between approximate arithmetic and acquired mathematical skills. Because these young children have yet to receive formal math education, especially symbolic addition and subtraction, and also because the TEMA largely tests elementary concepts of number and counting at this age, finding a positive relationship between approximate arithmetic training and TEMA improvement would suggest a deeper connection between approximate arithmetic training and math performance. Practically, approximate arithmetic training could benefit children with minimal background in math, in many cases from low-income families, allowing them to catch up with their math skills as early as possible.

The results of the current study demonstrate that repeated training in nonverbal quantity manipulation yields selective improvements in math ability in preschool children. The TEMA measures young children's formal and informal math abilities broadly, including numeral literacy, number facts, enumeration, and counting. Our analyses suggest that the effect of approximate arithmetic training was not selective to a particular type of early math problem—in particular, mental arithmetic—and instead suggest that there is a deep cognitive overlap between nonverbal quantity manipulation and learned math in general. Future research could attempt to isolate the mechanism by which approximate arithmetic training benefits learned mathematics by more systematically sampling different types of math skills before and after approximate arithmetic training.

An important question is whether the current results can be explained by placebo effects (e.g., see Boot, Simons, Stothart, & Stutts, 2013). Although the current study did not specifically test for a placebo effect, our previous work with adult participants makes that interpretation unlikely. Specifically, adults trained with (a) an approximate number comparison and approximate numerical matching task or (b) a numeral ordering task showed negligible benefits in math improvement (Park & Brannon, 2013, 2014). In fact, a recent study directly probed whether children's expectations about their own subsequent symbolic arithmetic performance differed as a function of different types of cognitive training tasks. The study reported that 6- to 8-year-olds did not show differences in their expectation between training that involved non-symbolic numerical quantity and training that did not involve numerical quantity (Dillon, Pires, Hyde, & Spelke, 2015).

Although our study provides an important first step that suggests that targeting approximate arithmetic in preschoolers could be an effective tool for improving math readiness, an important caveat is that assessing the feasibility and effectiveness of approximate arithmetic training in educational settings will require a much larger scale study with complete random assignment and experimenter-blind testing. Future research should compare approximate arithmetic training with other conditions that receive other types of math training designed to directly improve the targeted skills (e.g., teaching concepts and knowledge covered in the TEMA to improve the test). Similarly, future research should assess whether there are dose-dependent effects of the training and the longevity of the benefits for math. Finally, another important caveat to early educational interventions in general is that the effectiveness of such interventions diminishes over time, potentially suggesting that later academic success

is driven primarily by stable underlying characteristics related to learning (e.g., domain-general cognitive abilities, motivation, socioeconomic status) rather than by prior knowledge (Bailey, Watts, Littlefield, & Geary, 2014). Thus, from a practical point of view, an important future goal is to find ways not only to boost mathematical knowledge and skills but also to positively influence other latent sources of learning.

In sum, these findings provide additional evidence for a causal link between approximate non-symbolic numerical abilities and learned mathematics and may hold important practical implications for math education in preschool-aged children. Most existing research-based math intervention programs for early childhood focus directly on the skills that are targets for improvements (e.g., providing counting activities to improve counting abilities; but see Siegler & Ramani, 2009; for a review, see Clements & Sarama, 2011). Although these existing approaches are very valuable, our approach suggests that including training that taps deeper cognitive links may open new promising avenues for math intervention. In particular, a major advantage of approximate arithmetic training is that it can be used with very young children before they have mastered the meaning of number words and symbols, thereby providing opportunities to develop maximally effective ways of closing the achievement gap before kindergarten.

Acknowledgments

We are grateful to the child participants, their families, and the preschool administrators and teachers for making this study possible. We thank our Duke University Bass Connections team of undergraduates for their assistance in data collection and analysis. We appreciate Mark Paris's dedicated work in developing the two iPad games. This study was supported by NIH R01 HD079106 and a grant from the Initiative on Education and Human Development at Duke to E.M.B.

References

- Agrillo, C., Cohen Kadosh, R., & Dowker, A. (2014). *Numerical and arithmetic abilities in non-primate species*. Oxford, UK: Oxford University Press.
- Aubrey, C., Godfrey, R., & Dahl, S. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal*, 18(1), 27–46.
- Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological Science*, 25, 2017–2026.
- Baroody, A. J., Eiland, M., & Thompson, B. (2009). Fostering at-risk preschoolers' number sense. *Early Education and Development*, 20(1), 80–128.
- Beran, M. J., Perdue, B. M., & Evans, T. A. (2015). Monkey mathematical abilities. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 237). Oxford, UK: Oxford University Press.
- Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, 8, 445–454.
- Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children*, 78(1), 7–23.
- Chard, D. J., Baker, S. K., Clarke, B., Jungjohann, K., Davis, K., & Smolkowski, K. (2008). Preventing early mathematics difficulties: The feasibility of a rigorous kindergarten mathematics curriculum. *Learning Disability Quarterly*, 31(1), 11–20.
- Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education*, 38, 136–163.
- Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. *Science*, 333, 968–970.
- Dehaene, S. (1999). *The number sense: How the mind creates mathematics*. Oxford, UK: Oxford University Press.
- Dillon, M. R., Pires, A. C., Hyde, D. C., & Spelke, E. S. (2015). Children's expectations about training the approximate number system. *British Journal of Developmental Psychology*, 33, 411–418.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology*, 43, 1428–1446.
- Dunn, L. M., & Dunn, D. M. (1997). *Peabody Picture Vocabulary Test-Revised*. Minneapolis, MN: American Guidance Service.
- Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities*, 46, 166–181.
- 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.
- Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*, 97, 493–513.
- 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.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44, 43–74.

- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE*, *8*(1), e54651.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like day–night test. *Cognition*, *53*, 129–153.
- 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.
- Gilmore, C. K., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., ... Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE*, *8*(6), e67374.
- Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability* (3rd ed.). Austin, TX: Pro-Ed.
- Gobel, S. M., Watson, S. E., 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.
- Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*, 496–499.
- Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*, 1457–1465.
- 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.
- Halberda, J., Mazocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668.
- Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*, 92–107.
- Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review*, *18*, 1222–1229.
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 10382–10385.
- Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, *20*(2), 82–88.
- Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice*, *22*(1), 36–46.
- Jordan, N. C., Kaplan, D., Nabors Oláh, L., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, *77*, 153–175.
- Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*, 850–867.
- Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher "math talk". *Developmental Psychology*, *42*, 59–69.
- Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction*, *25*, 95–103.
- 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.
- Lindskog, M., & Winman, A. (2016). No evidence of learning in non-symbolic numerical tasks—A comment on Park and Brannon (2014). *Cognition*, *150*, 243–247.
- Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*, 256–261.
- Martin, R. B., Cirino, P. T., Sharp, C., & Barnes, M. (2014). Number and counting skills in kindergarten as predictors of Grade 1 mathematical skills. *Learning and Individual Differences*, *34*, 12–23.
- Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*(9), e23749.
- 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*(7), e67918.
- Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, *24*, 2013–2019.
- Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition*, *133*, 188–200.
- Park, J., & Brannon, E. M. (2016). How to interpret cognitive training studies: A reply to Lindskog & Winman. *Cognition*, *150*, 247–251.
- Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological Science*, *24*, 1037–1043.
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*, 499–503.
- Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*, 375–394.
- Sasanguie, D., Gobel, 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.
- Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*, 545–560.
- Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, *19*, 99–120.
- Tibber, M. S., Manasseh, G. S., 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.

- 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. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, B1–B11.
- Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63.