Published in final edited form as:

Cognition. 2021 February ; 207: 104521. doi:10.1016/j.cognition.2020.104521.

Failure to replicate the benefit of approximate arithmetic training for symbolic arithmetic fluency in adults

Emily Szkudlarek[a,*,1], **Joonkoo Park**[b,c], **Elizabeth M. Brannon**[a]

[a]University of Pennsylvania, Department of Psychology, 425 S. University Ave, Philadelphia, PA 19104, USA

[b]University of Massachusetts Amherst, Department of Psychological and Brain Sciences, 135 Hicks Way, Amherst, MA 01003, USA

[c]Commonwealth Honors College, University of Massachusetts Amherst, USA

Abstract

Previous research reported that college students' symbolic addition and subtraction fluency improved after training with non-symbolic, approximate addition and subtraction. These findings were widely interpreted as strong support for the hypothesis that the Approximate Number System (ANS) plays a causal role in symbolic mathematics, and that this relation holds into adulthood. Here we report four experiments that fail to find evidence for this causal relation. Experiment 1 examined whether the approximate arithmetic training effect exists within a shorter training period than originally reported (2 vs 6 days of training). Experiment 2 attempted to replicate and compare the approximate arithmetic training effect to a control training condition matched in working memory load. Experiments 3 and 4 replicated the original approximate arithmetic training experiments with a larger sample size. Across all four experiments ($N = 318$) approximate arithmetic training was no more effective at improving the arithmetic fluency of adults than training with control tasks. Results call into question any causal relationship between approximate, non-symbolic arithmetic and precise symbolic arithmetic.

Keywords

Numerical cognition; Approximate number system; Cognitive training; Approximate arithmetic; Math; Replication

*Corresponding author at: Wisconsin Center for Education Research, University of Wisconsin-Madison, 1025 West Johnson St., Room 698, Madison, WI 53706, USA. szkudlarek@wisc.edu (E. Szkudlarek).

[1]The present affiliation of Emily Szkudlarek is the University of Wisconsin-Madison.

Declaration of Competing Interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2020.104521.

1. Introduction

The Approximate Number System (ANS) supports the ability to represent, compare, and manipulate non-symbolic and approximate quantities without number symbols or language. The ANS is evident in diverse species of animals and early in human development (Feigenson, Dehaene, & Spelke, 2004). The number sense theory posits that the ANS is foundational to the development of symbolic mathematical knowledge (Dehaene, 2011; Gallistel & Gelman, 1992). A strong version of the number sense theory states that the ANS remains foundational for mathematical ability throughout the human lifespan (e.g., Bugden, DeWind, & Brannon, 2016; Feigenson, Libertus, & Halberda, 2013).

A large body of work relates the ANS to symbolic mathematical skill. The same brain regions that support symbolic math representations are recruited when children and rhesus macaques make approximate number judgments, supporting the neuronal recycling hypothesis (Dehaene & Cohen, 2007; Dehaene, Piazza, Pinel, & Cohen, 2003). There is a small, but significant, correlation between ANS acuity and a variety of symbolic math skills in both children and adults (for meta-analyses see Chen & Li, 2014; Fazio, Bailey, Thompson, & Siegler, 2014; Schneider et al., 2016). ANS acuity is longitudinally predictive of later math abilities (He et al., 2016; Soto-Calvo, Simmons, Willis, & Adams, 2015; Toll, Van Viersen, Kroesbergen, & Van Luit, 2015). Furthermore, some children with developmental dyscalculia have lower ANS acuity than age matched or skill matched control groups (Olsson, Östergren, & Träff, 2016; Piazza et al., 2010). These correlational findings suggest, but do not demonstrate, a causal link between ANS acuity and symbolic math ability throughout the lifespan.

The strong version of the ANS theory led to the prediction that there would be a causal link between tasks that engage the ANS and symbolic math performance, even among adult participants. In the first such experiment, Park and Brannon (2013) found that adults who trained with non-symbolic, approximate arithmetic problems answered more double and triple digit addition and subtraction problems correctly at post-test compared to pretest than subjects who trained with a numeral ordering task, a knowledge training task, or a no-contact control group. Approximate arithmetic training involved addition and subtraction of dot arrays over six or ten 25-min sessions. A subsequent experiment (Park & Brannon, 2014) replicated this effect, and found that non-symbolic, approximate arithmetic training improved symbolic arithmetic fluency more than training with a non-symbolic number comparison task, a visuo-spatial short-term memory task, or a numeral ordering training task. A separate research group replicated this effect with the finding that subjects trained on approximate arithmetic improved more on a symbolic arithmetic fluency test compared to subjects who spent the same amount of time answering general knowledge multiple choice questions (Au, Jaeggi, & Buschkuehl, 2018). Taken together, this work suggested that practice with non-symbolic addition and subtraction increases the ability to solve symbolic addition and subtraction problems.

Results of these training studies were interpreted as evidence for the strong version of the number sense theory that the ANS remains foundational for symbolic mathematics throughout the lifespan (e.g., Bugden et al., 2016; Feigenson et al., 2013; Hyde, Berteletti, &

Mou, 2016; Spelke, 2017). However, a caveat to this interpretation is that although the approximate arithmetic task requires the representation of large, approximate quantities, there is no evidence that the task changes ANS acuity. ANS acuity is quantified with a comparison task where participants identify which of two arrays of dots is greater in quantity, and the Weber fraction provides an estimate of the ratio between the two dot arrays that a participant requires for successful discrimination. Approximate arithmetic performance shows ratio dependence, and clearly requires participants to use their approximate sense of quantity, but the task also involves additional cognitive skills such as visual working memory and mental manipulation. An intuitive explanation for positive transfer between non-symbolic approximate arithmetic and symbolic arithmetic is that non-symbolic, approximate arithmetic training improves ANS acuity, which in turn leads to better symbolic math performance. However, contrary to this explanation, all current evidence indicates that non-symbolic arithmetic training does not change participant's ANS acuity (Au et al., 2018; Park & Brannon, 2014). Moreover, while ANS acuity is somewhat malleable with training (DeWind & Brannon, 2012) even extended training with non-symbolic dot comparison was not sufficient to improve adult symbolic arithmetic performance (Cochrane, Cui, Hubbard, & Green, 2018). Thus, any improvements in symbolic arithmetic fluency that resulted from approximate arithmetic training were not due to sharpening ANS acuity, but instead some other aspect of the non-symbolic, approximate arithmetic task. Park and Brannon (2014) instead proposed that the repeated manipulation of mental representations of quantity in arithmetic operations was the mechanism of symbolic arithmetic improvement.

The initial aim of the current set of experiments was to investigate the necessary parameters for transfer from approximate arithmetic training to symbolic arithmetic calculation to shed light on potential mechanisms of the approximate arithmetic effect. Experiment 1 asked whether the approximate arithmetic training effect exists with a shorter training period than originally reported (2 vs 6 days of training). Experiment 2 compared approximate arithmetic training to training matched in working memory load and the numerical quantities manipulated, but without addition or subtraction operations. Unexpectedly, we failed to replicate the approximate arithmetic training effect in Experiments 1 and 2. Consequently, we attempted more precise replications of Park & Brannon, 2013, 2014 in Experiment 3 (approximate arithmetic training compared to numeral ordering training) and a preregistered replication in Experiment 4 (approximate arithmetic training compared to approximate number comparison training). Finally, to increase the power for detecting any transfer effect we analyzed the data from Experiments 1–4 combined with the original data from the Park and Brannon experiments. Over four experiments and 318 participants we found no support for the original claim that training adults on the approximate arithmetic task benefits symbolic arithmetic fluency.

2. Methods

2.1. Subjects

Participants in Experiment 1 were recruited through the University of Pennsylvania′s Psychology Subject pool and received course credit for participation. Participants in

Experiments 2–4 were recruited with flyers distributed throughout the University of Pennsylvania's campus. These flyers advertised for participation in a "Brain Exercise" psychology experiment to study adult cognition. Subjects in Experiments 2 and 3 were paid in a lump sum at the completion of the post-test. Subjects in Experiment 4 were paid after each session. Flyers contained the same language as the recruitment flyers used in the original Park & Brannon studies (2013, 2014). Participants recruited through flyers were largely students at the University of Pennsylvania. Participants were required to speak English and be under the age of 35.

There were 38 subjects in Experiment 1, 78 subjects in Experiment 2, 91 subjects in Experiment 3, and 111 subjects in Experiment 4. A power analysis indicated 19 subjects in the approximate arithmetic training condition would achieve 80% power to detect the original effect size of improvement on the arithmetic fluency test as found in Experiment 1 of Park & Brannon, 2013 (the smallest original effect size found). We increased the number of subjects per condition in each subsequent experiment to increase the power to detect improvement on the arithmetic fluency test, resulting in our goal of 50 subjects per condition in Experiment 4 as reported in our preregistration. We collected 5 extra participants in the approximate arithmetic condition of Experiment 4 due a lower than anticipated rate of attrition. This sample size resulted in 99% power to detect the original effect size of improvement in Experiment 4 using a one sample two-sided *t*-test (cohen's d = 0.68, significance level = 0.05). Age and gender by experiment and training condition are reported in Table 1.

2.2. Procedure

For each experiment, participants were randomly assigned to conditions. All participants completed a pretest battery, 2 or 6 25-min training sessions, and a post-test battery that was matched to the pre-test battery. The pre and post-test battery in all experiments included an exact symbolic arithmetic test and various other tests (see Supplementary Methods and Table 1 for details). Experiment 1 consisted of only two sessions while Experiments 2–4 each consisted of 8 sessions. The training sessions were conducted on the same day as the pre and post-tests in Experiment 1 and on separate days for Experiments 2–4. Training sessions were 25 min in duration for all four experiments. All testing and training sessions took place in a quiet testing room with six computers. The average number of days between pre and posttest for participants in the approximate arithmetic training condition of each experiment is reported in Table 1. The order in which participants completed the pre and posttests was counterbalanced.

2.3. Training conditions

All four experiments consisted of an approximate arithmetic condition that was compared in a between subject design with at least one other training condition described below (Figure 1). All experimental tasks are available for download at https://osf.io/9e5ca/?view_only=9c2bd833cd9641d0af3c11c799cb4de7.

2.3.1. Approximate arithmetic (all experiments)—

The approximate arithmetic training condition was taken directly from the Park and Brannon studies (2013, 2014).

Participants mentally added or subtracted dot arrays ranging from 9 to 36 as the arrays moved behind or out from an occluder in the center of a computer monitor. Participants were then required to either compare the imagined sum or difference to a new target array, or to match the imagined sum or difference to one of two target arrays. In Experiments 1–3 participants responded by touching a touch screen monitor to indicate their choice. In Experiment 4, participants used the mouse to indicate their response. Dot arrays that represented the addends in the problem were visible for 1000 ms. Target dot arrays to be compared or matched were visible for 1500 ms before they were hidden behind a black circle. Both the matching and comparison trial types, and the addition and subtraction trial types were intermixed within 10 blocks of 20 trials each for each training session. Feedback was provided after each trial. The difficulty of the task was titrated to performance by decreasing the ratio between the imagined sum or difference and the target dot array. The numerical distance between the sum or difference and the target array varied in a log-base2 scale, the log difference level. All participants started training with a log difference level of 1.5. A log difference level of 1.5 is equivalent to a ratio of 2.83 (2^1.5 to 1) between the sum or difference and the alternative target array. For each 20-trial block, if performance was greater than 85% the log difference level decreased by one of the following randomly chosen values [0.13, 0.14, 0.15, 0.16, 0.17]. If performance over a block of 20 trials was less than 70% the log difference level increased by one of the values randomly chosen from the following set [0.08, 0.09, 0.10, 0.11, 0.12]. The log difference level achieved at the end of a training session was carried over into the next session.

2.3.2. Numeral symbol ordering (Experiments 1–3)—This training condition was taken directly from the Number Symbol Ordering training condition used in (Park & Brannon, 2013; Park & Brannon, 2014). Participants were required to reorder sets of three Arabic numerals before the numerals moved off of the screen by tapping the numerals, which rearranged themselves randomly with each touch. A maximum of three triads appeared on the screen at the same time. If the triad was moving to the left on the screen the numerals needed to be in ascending order. If the triad was moving to the right on the screen the numerals needed to be in descending order. Participants received feedback on whether the triad was in the right order as it entered a gray block at the edge of the screen. This gray block turned green if the triad was in the correct order, or red if the triad was in the incorrect order at the end of each trial. Task difficulty was titrated by varying the speed in which the triads travelled across the screen. Triad speed started at 125 pixels/s. If accuracy was greater than 90% over a 2.2 min span, the speed increased by one of the values chosen randomly from the following [10, 11, 12, 13, 14]. If accuracy was less than 80%, the speed decreased by one of the following values [4, 5, 6, 7, 8]. The speed at the end of one training session was carried over into the next session.

2.3.3. Approximate range (Experiment 2)—This novel training condition was designed to match the working memory load of the approximate arithmetic training condition and required mental manipulation of dot arrays without mental arithmetic. Participants in this training condition indicated whether or not the number of dots in a target dot array fell inside or outside the range of two previously viewed dot arrays. Participants watched as one dot array appeared in the middle of the screen for one second before it

moved behind either an occluder (a gray box) on the right or an occluder (an identical gray box) on the left, counterbalanced. A second dot array appeared for one second and moved behind the second occluder. On comparison trials, a target array appeared at the bottom of the screen for 1000 ms and participants touched the screen to indicate whether this target was within the range of the previous two arrays. The target array was hidden by a black circle after 1000 ms, but subjects had 3500 ms more to respond to match the approximate arithmetic training condition. For example, if the first array contained 30 dots and the second array contained 120 dots and the target array had 15 dots the correct choice was “outside” the range. On matching trials, two arrays appeared at the bottom of the screen and subjects chose the array that was within the range of the previous two animated arrays. For example, if the first array contained 30 dots and the second array contained 120 dots the correct choice would be 90 rather than 180. The number of dots in one array ranged from 4 to 256. The two arrays that defined a range always had a ratio of 1:4. The difficulty of this task was titrated by changing the ratio between one of the range defining arrays and the target dot array. The numerical distance between one of the range defining arrays and the target array varied in a log-base2 scale, the log difference level. All participants started training with a log difference level of 1. A log difference level of 1 is equivalent to a ratio of 2 (2^1 to 1) between one of the range defining arrays and the alternative target array. If performance over a block of 20 trials was greater than 85% the log difference level decreased by 0.10. If performance over a block of 20 trials was less than 70% the log difference level increased by 0.05. The log difference level achieved at the end of a training session was carried over into the next session.

2.3.4. Approximate number comparison (Experiment 4)—This training condition was taken directly from the Approximate Number Comparison training condition from Park and Brannon (2014). The task was to identify which of two dot arrays contained the greater number of dots. There were two trial types. In the mixed trial type, white and black dots appeared in an intermixed array on a gray background for 750 ms. Participants reported whether there were more black or white dots. In the other trial type participants saw two distinct black or white dot arrays on the screen at the same time, and chose which array was greater in numerosity. Participants responded with a mouse click. As in the approximate arithmetic training condition, difficulty was titrated using the log difference level. One of the dot arrays on each trial ranged from 16 to 32, and the other was determined by the log difference level. For example, if one array contained 16 dots, and the log difference level was 1.15, then the other array would contain either $16 \times 2^{1.15}$ or $16/2^{1.15}$ dots. The titration procedure was the same as used in the approximate arithmetic training condition.

2.4. Pre and post tests

2.4.1. Exact symbolic arithmetic test (all experiments)—The test of arithmetic fluency developed by Park and Brannon (2013, 2014) was used in all four of the current experiments. Participants solved two and three digit addition and subtraction problems over two five minute blocks. The operands of the problems ranged from 11 to 244. Problems were chosen randomly for each participant from a set of 800 potential problems for the pretest and a distinct set of 800 for the posttest (counterbalanced). The number of problems that required carrying and borrowing was matched for the two problem sets. Performance

was quantified as the total number of correct problems solved over the 10-min assessment. As a proxy measure of a test-retest reliability score, the average correlation between pre and posttest arithmetic fluency score across all experiments and conditions was 0.90, indicating high reliability (for a breakdown by experiment and training condition see Table S1).

2.4.2. Expectation matching questionnaire (all experiments)—To assess whether any differences in arithmetic fluency gains by condition were driven by differences in participant's expectation of improvement, we administered an expectation questionnaire (Dillon, Pires, Hyde, & Spelke, 2015). This questionnaire was administered after the completion of all other post-tests and is reproduced in the supplement. Questions included "After playing this number symbol/dot game, do you think you would answer arithmetic questions more quickly?" and "After playing this number symbol/dot game, do you think you would get more arithmetic questions correct?". Participants indicated their response from 1 (strongly disagree) to 10 (strongly agree).

2.4.3. Additional pre and post-test assessments—Participants in each experiment completed two to four additional pre and posttests. These assessments were originally included as control tasks to assess whether any improvements in symbolic arithmetic after non-symbolic approximate arithmetic training were specific to symbolic arithmetic fluency. Given that the results were null for our outcome of interest, we restrict the description of the control assessments and the accompanying data to Table 1 and the Supplementary Material.

3. Theory and calculation

3.1. Analysis plan

Gain scores for each participant were calculated for the exact symbolic arithmetic assessment, by subtracting the number of problems solved correctly in 10 min at pos*t*-test from the number of problems solved correctly in 10 min at pretest. We removed any arithmetic gain score that was smaller than Q1 − × IQR or larger than Q3 + 2 × IQR, where Q1 is the first quartile, Q3 is the third quartile, and IQR is the interquartile range. In this analysis, quartiles were calculated with the data from each experiment separately. We then conducted a two-sample t-test or a one-way ANOVA to examine whether there was a significant difference between the average arithmetic fluency gain score by training condition. The transfer effect from training condition to arithmetic fluency score was further assessed using an analysis of covariance (ANCOVA) where pretest arithmetic fluency and training condition were used to predict post-test arithmetic fluency score. These analyses were preregistered for Experiment 4 with asPredicted.org (http://aspredicted.org/blind.php?x=kh6sy2). As a complement to the frequentist analysis of the training effect, we also report a Bayesian analysis of this effect for each experiment to examine the relative support for both our hypothesis of interest and the null hypothesis. We conducted a Bayesian t-test or ANOVA, dependent on the number of training conditions for each experiment. We set a non-informative Jeffreys prior width of 0.5 to correspond to a small effect (Morey & Rouder, 2011). These analyses result in a Bayes factor (BF_{10}), which can be interpreted as the likelihood ratio for the alternative hypothesis over the null. Given that the Bayes factor (BF_{10}) is a ratio of the likelihood for the alternative hypothesis over the null hypothesis, the

inverse of the Bayes factor (BF_{01}) can be interpreted as the likelihood ratio for evidence of the null hypothesis over the alternative hypothesis. Following Jeffreys (1961) we use the following designations to interpret the strength of the Bayes factors: 0–3 offer anecdotal support for H1, 3–10 moderate support for the H1, 10–30 strong support for H1, 30–100 very strong evidence for H1, and values greater than 100 offer decisive evidence for H1. We use the inverse of these ranges to interpret support for the null hypothesis (BF_{01} anecdotal 0.33–0, moderate 0.10–0.33, strong 0.10–0.03, very strong 0.03–0.01) To facilitate comparison with the present data we conducted a Bayesian analysis of Park and Brannon's (2013, 2014) previously reported data broken down by experiment.

Finally, we combined the data from all experiments, including the data from (Park & Brannon, 2013; Park & Brannon, 2014) to test the hypothesis that approximate arithmetic training improves exact symbolic arithmetic fluency more than any of the alternative training conditions: approximate number comparison, visuo-spatial short term memory, numerical symbol ordering, approximate range, knowledge training, and a no contact control. This combined analysis is possible because the same pre and post arithmetic fluency test was used across all experiments and training conditions. We again used complementary frequentist and Bayesian approaches. We first report a one-way ANOVA with training condition as a factor and exact symbolic arithmetic gain score as the outcome. Then, we report a one-way Bayesian ANOVA testing whether condition as a factor adds significant variance over the model with the mean intercept only. This analysis tests whether any of the training conditions create significant differences in arithmetic fluency gain, however, we had a specific hypothesis that approximate arithmetic training improves arithmetic fluency more than any of the other training conditions. To examine this specific hypothesis, we also report a contrast between the average arithmetic fluency gain score for the approximate arithmetic condition compared to all other conditions. We followed up this analysis with one sided *t*-tests between the average gain score for participants in the approximate arithmetic training condition and each other training condition. Finally, we compared the effect size of the arithmetic fluency gain found in the non-symbolic, approximate arithmetic condition across all seven experiments to test the robustness of the transfer effect regardless of the control conditions.

4. Results

4.1. Analysis of approximate arithmetic training performance

To quantify training performance, we calculated the mean log difference score of the matching and comparison trial types at the end of each training session for each participant in the approximate arithmetic training condition to match the analysis in the original Park and Brannon experiments. A one-way ANOVA indicated there was no significant difference between the last log difference level reached on training day 6 by experiment across the three Park and Brannon studies and the current experiments 2–4 ($F_{5,182} = 1.32$, $p = .26$, η_p^2; Fig. S7). This finding indicates that improvement in participants' ability to add and subtract dot arrays over the course of training was consistent across the original and current experiments. There were also no significant differences between mean log difference level by experiment on training session 2 for Experiments 1–4 and the three experiments

conducted by Park and Brannon ($F_{6,202} = 0.85$, $p = .53$, η_P^2). These analyses suggest that motivation to complete the training task in the approximate arithmetic training condition was similar in the current experiments and the prior studies by Park and Brannon.

4.2. Experiment 1

A two-sample *t*-test indicated no significant differences in mean number of arithmetic problems solved correctly at pretest by condition (63 vs 73 problems; $t_{36} = -1.20$, $p = .24$, d = −0.39 95% CI [−1.1 0.27]) suggesting that random assignment was effective. There was no significant difference between the average arithmetic gain score for the approximate arithmetic and numeral ordering training conditions (Figs. 2 & 3; 4.3 vs 10 problems; $t_{36} = -1.54$, $p = .13$, d = −0.50 95% CI [−1.2 0.17]). An ANCOVA confirmed no significant effect of condition on post-test arithmetic fluency score when controlling for pretest arithmetic fluency score ($F_{1,35} = 2.31$, $p = .14$, η_P^2) significant pretest score by condition interaction ($F_{1,34} = 0.003$, $p = .96$). The complementary Bayesian t-test indicated a $BF_{10} = 0.98$. A Bayes factor close to 1 suggests no evidence for either the alternative or the null hypothesis.

4.3. Experiment 2

One participant in the approximate range condition was removed from the sample due to an outlier gain score. There were no significant differences in pretest score by condition indicating that random assignment was effective (Approximate Arithmetic 67 problems, Numeral Ordering 68 problems, Approximate Range 70 problems; $F_{2,74} = 0.136$, $p = .87$, η_P^2). A one-way ANOVA with training condition as a factor indicated no differences in arithmetic gain score by training condition (Figs. 2 & 3; Approximate Arithmetic 5.2 problems, Numeral Ordering 7.3 problems, Approximate Range 5.5 problems; $F_{2,74} = 0.257$, $p = .77$, η_P^2). An ANCOVA indicated no significant effect of condition on pos*t*-test arithmetic fluency score when controlling for pretest scores ($F_{2,73} = 0.265$, p = .77, η_P^2), and no significant pretest by condition interaction ($F_{2,71} = 0.661$, $p = .52$). A Bayesian ANOVA resulted in a $BF_{10} = 0.137$, suggesting moderate evidence for the null hypothesis of no difference in arithmetic fluency gain score by condition.

4.4. Experiment 3

Two participants in the numeral ordering training condition were removed due to outlier gain scores. There was no significant difference in pretest arithmetic scores by condition indicating that random assignment was effective (75 vs 70 problems; $t_{87} = 0.873$, $p = .38$, d =0.19 95% CI [−0.24 0.61])). Again, there was no significant difference in arithmetic gain score by condition (Figs. 2 & 3; 8.0 vs 6.1 problems; $t_{87} = 0.873$, p = .38, d = 0.19 95% CI [−0.24 0.61]). An ANCOVA confirmed no significant effect of condition on pos*t*-test arithmetic fluency score when controlling for arithmetic fluency pretest score ($F_{1,86} = 0.681$, $p = .41$, η_P^2), and no significant pretest by condition interaction ($F_{1,85} = 3.14$, $p = .08$). The Bayesian t-test indicated a $BF_{10} = 0.60$, indicating anecdotal support for the null hypothesis that the difference between the mean gain scores of each condition is zero.

4.5. Experiment 4

There was a marginal difference between the arithmetic fluency scores at pretest by condition (58 vs 70 problems; $t_{109} = -1.92$, $p = .06$, d = −0.37 95% CI [−0.75 0.01]). This effect was driven by the pretest score of one subject in the approximate number comparison condition who scored over 6 standard deviations above the mean pretest score. With this outlier pretest score removed, there was no longer a marginal difference in pretest score by condition (58 vs 66 problems; $t_{108} = -1.66$, $p = .10$, d = −0.32 95% CI [−0.70 0.06]). However, this participant was not removed from our subsequent analyses because their gain score was within our outlier cutoffs.[2] There was no significant difference between arithmetic gain score by training condition (Figs. 2 & 3; 7.1 vs 10 problems; $t_{109} = -1.35$, $p = .18$, d = −0.26 95% CI [−0.63 0.12]).). An ANCOVA confirmed no significant effect of condition on arithmetic fluency post-test score when controlling for pretest score ($F_{1,108} = 2.16$, $p = .14$, η_P^2), and no significant condition by pretest score interaction ($F_{1,107} = 0.261$, $p = .61$). The Bayesian t-test indicated a $BF_{10} = 0.85$, indicating anecdotal support for the null hypothesis that there is no difference in gain score by condition.

4.6. Bayesian re-analysis of Park & Brannon, 2013, 2014

4.6.1. Park & Brannon, 2013 Experiment 1—A Bayesian t-test yielded a $BF_{10} = 3.01$, suggesting moderate support for the alternative hypothesis of a significant difference between the arithmetic fluency gain scores for the approximate arithmetic (9.3 problems) and no contact control groups (0.31 problems). This reanalysis is consistent with the conclusions reported in Park & Brannon, 2013.

4.6.2. Park & Brannon, 2013 Experiment 2—A Bayesian ANOVA indicated a $BF_{10} = 1.71$, indicating anecdotal evidence for the alternative hypothesis of a significant difference in arithmetic fluency gain score by training condition (Approximate Arithmetic 15 problems, Numeral Ordering 5.1 problems, Knowledge Training 6.1 problems). This reanalysis is consistent with the conclusions reported in Park & Brannon, 2013.

4.6.3. Park & Brannon, 2014 Experiment 1—A Bayesian ANOVA indicated $BF_{10} = 4.28$, indicating moderate evidence for the alternative hypothesis of a significant difference in arithmetic fluency gain score by training condition (Approximate Arithmetic 14 problems, Numeral Ordering 5 problems, Approximate Number Comparison −2.9 problems, Visuo-spatial Short Term Memory 4.4). This reanalysis is consistent with the conclusions reported in Park & Brannon, 2014.

4.7. Combined analysis

We combined all the data from the current experiments and the three previous experiments conducted by Park and Brannon (2013, 2014) to yield a dataset with 486 individual arithmetic fluency gain scores across seven training conditions. A new outlier analysis with the full data set resulted in four arithmetic gain score outliers across all seven experiments:

[2]Removal of this subject does not change the significance of any analyses. Difference between gain scores by training condition ($t_{108} = -1.45$, $p = .15$; ANCOVA $F_{1,107} = 2.36$, $p = .13$, η_P^2; $BF_{10} = 0.92$).

one from Park & Brannon, 2013 Experiment 2, one from Park & Brannon, 2014 Experiment 1, one from Experiment 2, and one from Experiment 4. These participants were excluded from the current analysis. A one-way ANOVA predicting pretest arithmetic fluency score by condition was not significant (Fig. S5; $F_{6,475} = 0.855$, $p = .53$, η^2_P) indicating there were no significant differences in arithmetic fluency score by condition at pretest. Moreover, a one-way ANOVA predicting arithmetic fluency pretest score by experiment was also not significant (Fig. S5; $F_{6,475} = 1.27$, $p = .27$, η^2_P) suggesting that the samples for each experiment had comparable initial arithmetic performance. Crucial to our main hypothesis, a one-way ANOVA predicting arithmetic fluency gain score with condition as a factor indicated no significant differences by condition (Fig. 4; $F_{6,475} = 1.69$, $p = .12$, $\eta^2_P = 0.02$). An ANCOVA confirmed no significant effect of condition on arithmetic fluency post-test score when controlling for pretest score ($F_{6,474} = 1.61$, $p = .14$, η^2_P), and no significant condition by pretest score interaction ($F_{6,468} = 0.718$, $p = .64$). The complementary Bayesian one-way ANOVA resulted in a $BF_{10} = 0.14$ for the condition factor. This provides moderate evidence that the model with only the mean intercept is a better model of arithmetic gain score than a model with training condition as a factor. The current data is seven times (i.e., 1/0.14 = 7.14) more likely to occur under the null hypothesis that the intercept only model is a better model of the data than the alternative model that training condition explains variance in arithmetic fluency gain score.

To test the specific hypothesis that approximate arithmetic training improves arithmetic fluency more than any other training condition, we ran a contrast between the approximate arithmetic condition and all other conditions. This ANOVA indicated a significant difference between the approximate arithmetic condition and the other training conditions as a whole (Fig. 4; $F_{1,480} = 3.93$, $p = .048$, $BF_{10} = 0.68$). However, one sided *t*-tests between the average gain score for the approximate arithmetic training condition and every other condition revealed that this effect was driven by greater arithmetic fluency gain scores in the approximate arithmetic training condition compared to the no contact control condition ($t_{232} = 3.12$, $p = .001$, $BF_{10} = 16.0$). None of the other one-tailed t-tests comparing arithmetic fluency gain scores for the approximate arithmetic training and each of the other training conditions were significant (numeral ordering $t_{325} = 0.987$, $p = .16$, $BF_{10} = 0.27$; approximate number comparison $t_{279} = 0.844$, $p = .20$, $BF_{10} = 0.28$; approximate range $t_{229} = 0.986$, p = .16, $BF_{10} = 0.44$; visuo-spatial short term memory $t_{224} = 1.20$, p = .12, $BF_{10} = 0.56$; knowledge training $t_{221} = 0.612$, p = .27, $BF_{10} = 0.41$).

4.8. Effect size of approximate arithmetic training improvement in arithmetic fluency

Finally, we compared the effect size of the gain scores within the non-symbolic, approximate arithmetic training condition across experiments with 6 days of training (Experiments 2–4, Experiment 2 in Park & Brannon, 2013, Experiment 1 in Park & Brannon, 2014). The original experiments reported effect sizes for the approximate arithmetic training condition of d = 1.08 (15.4 problems, Park & Brannon, 2013 Experiment 2), and d = 1.10 (14.4 problems, Park & Brannon, 2014 Experiment 1). The effect sizes for Experiments 2–4 of the present study were d = 0.45. d = 0.82, and d = 0.54, corresponding to an increase of 5.19, 8.00, and 7.05 problems answered correctly respectively. A one-way

ANOVA testing for a significant difference in arithmetic fluency gain score by experiment revealed a significant difference (Fig. 4; $F_{4,159} = 3.10$ $p = .02$, $\eta_P^2 = 0.07$). Pairwise tests revealed that the effect size for Experiments 2–4 were smaller than the effect size found in Park & Brannon, 2013 Experiment 2 (Experiment 2 $p = .01$, Experiment 3 p = .04, Experiment 4 p = .02), and the effect sizes found in Experiments 2 and 4 were smaller than the effect size found in Park & Brannon, 2014 (Experiment 2 p = .01, Experiment 4 $p = .03$). However, none of these pairwise comparisons survived the Holm correction for multiple comparisons.

4.9. Expectation matching analysis

When combining all data from Experiments 1–4 there was no correlation between a participant's expectation of increased accuracy or speed on the arithmetic fluency test and a participant's actual improvement on the arithmetic fluency test (accuracy r (286) = −0.03, $p = .67$, 95% CI [−0.14 0.09]; speed r (286) = −0.06, $p = .33$, 95% CI [−0.17 0.06]). For a comparison of expectation for improvement by condition, please see the supplementary material.

5. Discussion

The original goal of the current study was to replicate the finding that non-symbolic approximate arithmetic training improves symbolic arithmetic fluency, and to build on this finding by probing the mechanism of the transfer effect. However, we were unable to replicate the original finding across four independent experiments. Bayesian analyses for each of the four current experiments provided weak to moderate evidence in favor of the null hypothesis of no significant difference in arithmetic fluency gain by training condition. To increase the power to detect an approximate arithmetic training effect, we combined the data from all four experiments and the original data from Experiments 1 and 2 of Park & Brannon, 2013, and Experiment 1 of Park & Brannon, 2014. However, even with the large sample size of 486 (209 in the approximate arithmetic training condition alone) there was no significant difference in arithmetic fluency gain score by training condition. A Bayesian analysis of this combined data set indicated that the data is seven times more likely under the null hypothesis of no difference between training conditions than under the alternative hypothesis where training condition is predictive of arithmetic fluency gain.

All training conditions did, on average, improve participants scores on the arithmetic fluency test. However, without a significant difference in training effect by condition, this increase in performance is likely due to a test-retest effect. The testing environment, test instructions, and the method of response were the same during the pre and post-test sessions, and this similarity could result in improved performance at post-test. At the same time, without a no contact control condition in each experiment, we cannot definitively claim that a test-retest effect accounts for improved performance. It is possible that our control training conditions improved arithmetic fluency to the same degree as approximate arithmetic training, and this across the board improvement could account for our lack of a difference between training conditions at post-test. However, we consider this explanation unlikely due to the variety of alternative training conditions tested across all experiments.

We did not find evidence that participants' expectations of improvement impacted participants' actual improvement on the arithmetic posttest. There was no significant correlation between a participant's expectation of improvement and their actual improvement. Moreover, if anything, participants in the numeral ordering training condition expected to improve their arithmetic score to a greater degree than participants in the other numerical training conditions, and yet their gains were no greater than the other training conditions. This lack of correspondence between expected gains and actual gains is unexpected, but may reflect the fact that regardless of a participant's expectation that 'brain training' should improve their cognitive ability, none of our training conditions were an effective way to improve arithmetic fluency.

We were also unable to replicate the effect size of the arithmetic fluency gain in the approximate arithmetic condition in the current experiments. The effect sizes for the approximate arithmetic condition in the current experiments were lower than the effect sizes found in Park & Brannon, 2013 Experiment 2 and Park & Brannon, 2014 Experiment 1. This may reflect the "winner's curse" and be a consequence of the small sample sizes used in the Park and Brannon studies (Button et al., 2013). An underpowered experiment will tend to overestimate the size of the effect because, by chance, the original experiment found an estimate of the effect that was large enough to pass our statistical threshold for significance ($p < .05$). Consequently, subsequent experiments that attempt to replicate this effect will tend to find smaller effect sizes that are closer to the actual true effect size (Button et al., 2013). Thus, the smaller "true" effect size for the approximate arithmetic condition found in the replication experiments is not significantly different than the effects measured for the control training conditions. We consider this the best explanation for our inability to replicate Park and Brannon (2013, 2014).

While the small sample size used in the original experiments may have led to false positive results and our inability to replicate, we also considered other potential reasons for our discrepant results. The Park and Brannon studies (2013, 2014) were run at a different university than the current experiments, however, these universities are well matched in terms of academic performance, entrance selectivity, and student demographics. We found no systematic differences in pretest arithmetic fluency scores or engagement with the approximate arithmetic training task. The same participant recruitment flyers and computer programs for both the arithmetic fluency outcome measure and the training tasks were used in the both the original and current experiments. We did identify one minor difference in the way the conditions were run. In the original experiments training cohorts tested in the same room were largely made up of one training condition to prevent participants seeing the other training conditions. In the current experiments, subjects were randomly assigned to conditions to reduce the possibility of cohort effects. Although we did not anticipate this minor difference to impact our results, recent studies suggest small differences in experimenter or participant knowledge of conditions can influence results (e.g., Gilder & Heerey, 2018). Additionally, separate training condition cohorts could result in an uneven distribution of motivation to complete a 'brain training' experiment because of the sequential nature of this method of testing. Participants may have different levels of motivation at different timepoints in the academic semester.

The results of the current experiments alone and in combination with the results of Park and Brannon (2013, 2014) lead us to conclude that there is no evidence that approximate arithmetic training improves arithmetic fluency in adults. However, it remains plausible that such training would be more effective in children. Previous studies with preschool and elementary school age children report that approximate arithmetic training has positive effects on tests of basic arithmetic calculation (Hyde, Khanum, & Spelke, 2014), symbolic number line placement (Khanum, Hanif, Spelke, Berteletti, & Hyde, 2016) and standardized math tests (Park, Bermudez, Roberts, & Brannon, 2016; Szkudlarek & Brannon, 2018). Interventions that incorporate an approximate arithmetic component into a larger intervention have also improved the symbolic math skill of children (Dillon, Kannan, Dean, Spelke, & Duflo, 2017; Käser et al., 2013; Obersteiner, Reiss, & Ufer, 2013; Sella, Tressoldi, Lucangeli, & Zorzi, 2016). This pattern of results would be consistent with the finding that the correlation between the ANS and symbolic mathematics is stronger when children are just beginning to learn their numbers, but weakens as children engage in formal math education (Fazio et al., 2014). Approximate arithmetic training may be more effective in young children because it improves their conceptual understanding of arithmetic. The same training may not benefit adults as basic arithmetic skills become algorithmic after years of formal math education.

What are the implications of our results for the number sense theory? With our inability to replicate Park and Brannon (2013, 2014) there is little evidence in the literature that can make a causal claim in support of the strong version of the number sense theory, which hypothesizes a causal link between the ANS and symbolic mathematics in adulthood. The majority of evidence supporting this theory remains correlational in nature, and specifically, relies on a significant, although weak relation between ANS acuity and symbolic math skills (Chen & Li, 2014; Schneider et al., 2016). Importantly, approximate arithmetic training did not change ANS acuity in our experiments or in prior studies (Au et al., 2018; Park & Brannon, 2014). Approximate arithmetic is not a direct measure of ANS acuity, but instead uses ANS representations *and* requires additional cognitive processes (e.g., visual working memory and mental manipulation). It remains possible that an intervention that could effectively induce a significant change in ANS acuity would in fact change symbolic arithmetic performance (Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Szűcs & Myers, 2017). Moreover, it is also possible that a different symbolic math skill other than arithmetic fluency could benefit from an ANS involved training, even among adults.

6. Conclusions

In sum, the present experiments lead us to reject the conclusion that training approximate arithmetic benefits symbolic arithmetic performance in adults. While the hypotheses that inspired these experiments emerged from the strong version of the number sense theory that the ANS is foundational for symbolic math throughout the lifespan, our current results no longer provide evidence for a causal link between the ANS and symbolic arithmetic in adult subjects. The educational implications of approximate arithmetic training in children remain unclear, but our results suggest that causal evidence for a link between non-symbolic and symbolic arithmetic is more likely to emerge from research with children at the beginning of their mathematical education.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Yvette Almaguer, Bonnie Zuckerman, I'Mani Sellers, Wendy Arce, Esther Adegbulugbe, Chung Chae, Lauren Paulson, Yoon Joo Kim, Zoe Belardo, Luis Rosario, Yvette Frimpong, Haobai Zhang, Joseph Dembik, Amanda Fields, Marilyn Baffoe-Bonnie, and Jiaqi Wu for their assistance with data collection. We would also like to thank Stephanie Bugden and Nicholas DeWind for helpful discussion of the manuscript. This research was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development NIH RO1 HD079106 to EMB and F31 HD095579-01 to ES.

References

Au J, Jaeggi SM, & Buschkuehl M (2018). Effects of non-symbolic arithmetic training on symbolic arithmetic and the approximate number system. Acta Psychologica, 185, 1–12. 10.1016/j.actpsy.2018.01.005. [PubMed: 29407240]

Bugden S, DeWind NK, & Brannon EM (2016). Using cognitive training studies to unravel the mechanisms by which the approximate number system supports symbolic math ability. Current Opinion in Behavioral Sciences. 10.1016/j.cobeha.2016.05.002.

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, & Munafò MR (2013). Power failure: Why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience, 14(5), 365–376. 10.1038/nrn3475. [PubMed: 23571845]

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. 10.1016/j.actpsy.2014.01.016. [PubMed: 24583622]

Cochrane A, Cui L, Hubbard EM, & Green CS (2018). "Approximate number system" training: A perceptual learning approach. Attention, Perception, & Psychophysics. 10.3758/s13414-018-01636-w.

Dehaene S, & Cohen L (2007). Cultural recycling of cortical maps. Neuron, 56(2), 384–398. 10.1016/j.neuron.2007.10.004. [PubMed: 17964253]

Dehaene S, Piazza M, Pinel P, & Cohen L (2003). Three parietal circuits for number processing. Cognitive Neuropsychology, 20(3–6), 487–506. 10.1080/02643290244000239. [PubMed: 20957581]

Dehaene S (2011). The number sense: How the mind creates mathematics. USA: OUP.

DeWind NK, & Brannon EM (2012). Malleability of the approximate number system: Effects of feedback and training. Frontiers in Human Neuroscience, 6 10.3389/fnhum.2012.00068.

Dillon MR, Kannan H, Dean JT, Spelke ES, & Duflo E (2017). Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics. Science, 357(6346), 47–55. [PubMed: 28684518]

Dillon MR, Pires AC, Hyde DC, & Spelke ES (2015). Children's expectations about training the approximate number system. British Journal of Developmental Psychology, 33, 411–418. 10.1111/bjdp.12118.

Fazio LK, Bailey DH, Thompson CA, & Siegler RS (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. Journal of Experimental Child Psychology, 123, 53–72. 10.1016/j.jecp.2014.01.013. [PubMed: 24699178]

Feigenson L, Dehaene S, & Spelke E (2004). Core systems of number. Trends in Cognitive Sciences, 8, 307–314. 10.1016/j.tics.2004.05.002. [PubMed: 15242690]

Feigenson L, Libertus ME, & Halberda J (2013). Links between the intuitive sense of number and formal mathematics ability. Child Development Perspectives, 7(2), 74–79. 10.1111/cdep.12019. [PubMed: 24443651]

Gallistel CR, & Gelman R (1992). Preverbal and verbal counting and computation. Cognition, 44, 43–74. [PubMed: 1511586]

Gilder TSE, & Heerey EA (2018). The role of experimenter belief in social priming. Psychological Science, 29, 403–417. 10.1177/0956797617737128. [PubMed: 29377787]

Halberda J, Ly R, Wilmer JB, Naiman DQ, & Germine L (2012). Number sense across the lifespan as revealed by a massive internet-based sample. Proceedings of the National Academy of Sciences, 109(28), 11116–11120. 10.1073/pnas.1200196109.

He Y, Zhou X, Shi D, Song H, Zhang H, & Shi J (2016). New evidence on causal relationship between approximate number system (ANS) acuity and arithmetic ability in elementary-school students: a longitudinal cross-lagged analysis. Frontiers in Psychology, 7 10.3389/fpsyg.2016.01052.

Hyde DC, Berteletti I, & Mou Y (2016). Approximate numerical abilities and mathematics In, Vol. 227 Progress in brain research (pp. 335–351). Elsevier http://linkinghub.elsevier.com/retrieve/pii/S0079612316300371. [PubMed: 27339018]

Hyde DC, Khanum S, & Spelke ES (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. Cognition, 131 (1), 92–107. 10.1016/j.cognition.2013.12.007. [PubMed: 24462713]

Jeffreys H (1961). Theory of probability (3, pp. 107–110). Oxford, UK: Oxford University Press.

Käser T, Baschera G-M, Kohn J, Kucian K, Richtmann V, Grond U, … von Aster M (2013). Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition. Frontiers in Psychology, 4 10.3389/fpsyg.2013.00489.

Khanum S, Hanif R, Spelke ES, Berteletti I, & Hyde DC (2016). Effects of non-symbolic approximate number practice on symbolic numerical abilities in Pakistani children. PLoS One, 11, Article e0164436.

Morey RD, & Rouder JN (2011). Bayes factor approaches for testing interval null hypotheses. Psychological Methods, 16(4), 406–419. 10.1037/a0024377. [PubMed: 21787084]

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. 10.1016/j.learninstruc.2012.08.004.

Olsson L, Östergren R, & Träff U (2016). Developmental dyscalculia: A deficit in the approximate number system or an access deficit? Cognitive Development, 39, 154–167. 10.1016/j.cogdev.2016.04.006.

Park J, & Brannon EM (2013). Training the approximate number system improves math proficiency. Psychological Science, 24, 2013–2019. 10.1177/0956797613482944. [PubMed: 23921769]

Park J, Bermudez V, Roberts RC, & Brannon EM (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. Journal of Experimental Child Psychology, 152, 278–293. 10.1016/j.jecp.2016.07.011. [PubMed: 27596808]

Park J, & Brannon EM (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. Cognition, 133, 188–200. 10.1016/j.cognition.2014.06.011. [PubMed: 25044247]

Piazza M, Facoetti A, Trussardi AN, 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. 10.1016/j.cognition.2010.03.012. [PubMed: 20381023]

Schneider M, Beeres K, Coban L, Merz S, Susan Schmidt S, Stricker J, & De Smedt B (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. Developmental Science, 20, Article e12372. 10.1111/desc.12372.

Sella F, Tressoldi P, Lucangeli D, & Zorzi M (2016). Training numerical skills with the adaptive videogame "The Number Race": A randomized controlled trial on preschoolers. Trends in Neuroscience and Education, 5, 20–29. 10.1016/j.tine.2016.02.002.

Soto-Calvo E, Simmons FR, Willis C, & Adams A-M (2015). Identifying the cognitive predictors of early counting and calculation skills: Evidence from a longitudinal study. Journal of Experimental Child Psychology, 140, 16–37. 10.1016/j.jecp.2015.06.011. [PubMed: 26218332]

Spelke ES (2017). Core knowledge, language, and number. Language Learning and Development, 13(2), 147–170. 10.1080/15475441.2016.1263572.

Szkudlarek E, & Brannon EM (2018). Approximate arithmetic training improves informal math performance in low achieving preschoolers. Frontiers in Psychology, 9 10.3389/fpsyg.2018.00606.

Szűcs D, & Myers T (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: A systematic review. Trends in Neuroscience and Education, 6, 187–203. 10.1016/j.tine.2016.11.002.

Toll SW, Van Viersen S, Kroesbergen EH, & Van Luit JE (2015). The development of (non-) symbolic comparison skills throughout kindergarten and their relations with basic mathematical skills. Learning and Individual Differences, 38, 10–17. http://www.sciencedirect.com/science/article/pii/S1041608015000023.

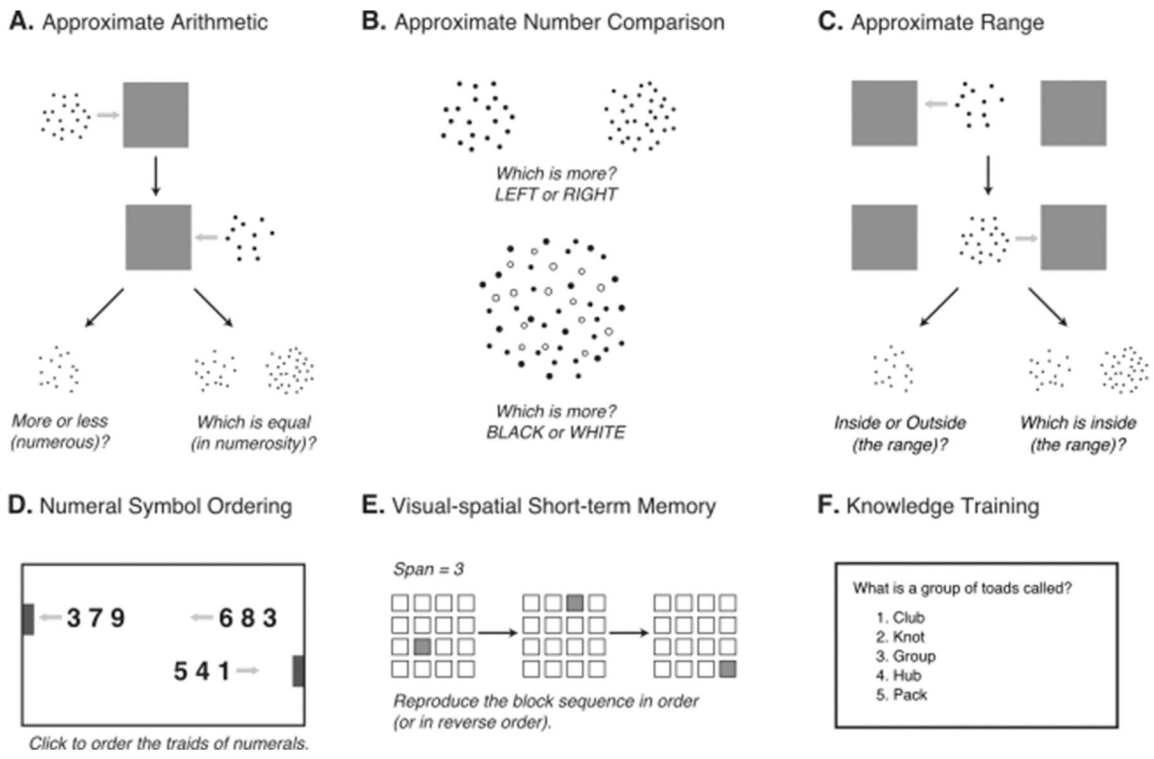

Fig. 1.

Training conditions used in the current experiments and the original Park and Brannon (2013, 2014) experiments. All experiments included A) the approximate arithmetic training condition and B–F) at least one control condition. Experiment 1 included A) approximate arithmetic training and D) the numeral symbol ordering training. Experiment 2 included A) approximate arithmetic training and C) approximate range training and D) numeral ordering training. Experiment 3 included A) approximate arithmetic training and D) numeral ordering training. Experiment 4 included A) approximate arithmetic training and B) approximate number comparison training. Park and Brannon (2013) included A) approximate arithmetic training, D) numeral symbolic ordering training and F) knowledge training. Park and Brannon (2014) included A) approximate arithmetic training, B) approximate number comparison training, D) numeral symbol ordering training, and E) visual-spatial short-term memory training.

Figure modified with permission from Park and Brannon (2014).

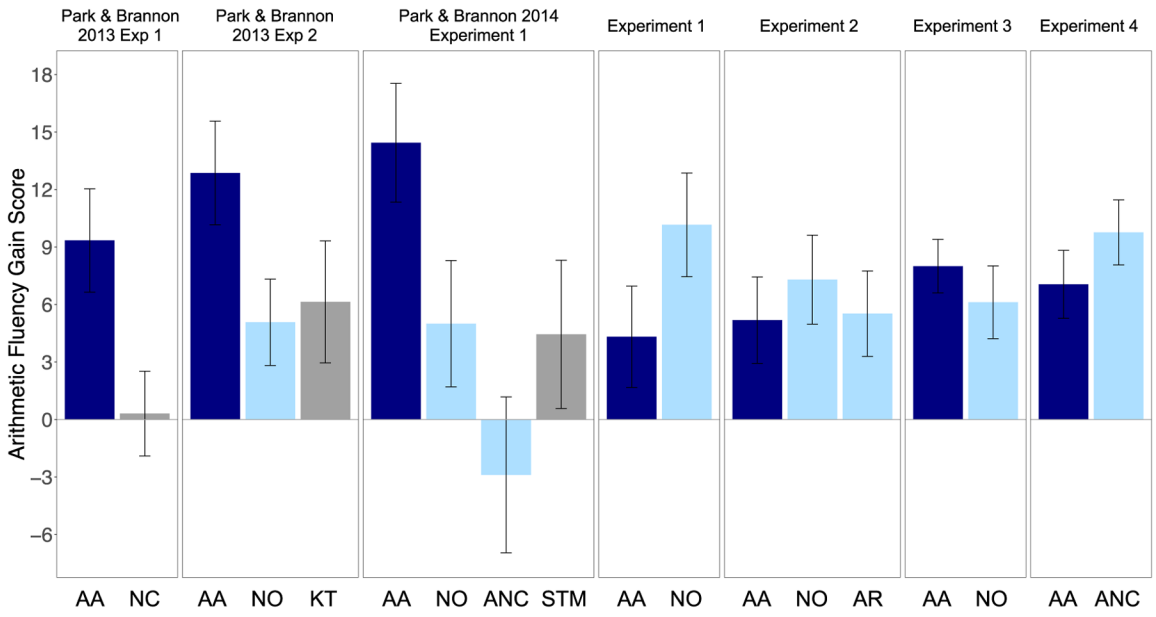

Fig. 2.

Bar plot of arithmetic fluency gain score by condition and experiment. AA = approximate arithmetic, NO = numeral symbol ordering, ANC = approximate number comparison, AR = approximate range, STM = visuo-spatial short term memory, KT = knowledge training, NC = no contact. The arithmetic fluency gain score is plotted in terms of the number of correct arithmetic questions (post-test minus pre-test). The bars colored in shades of blue are numerical training conditions. The gray bars represent non-numerical or no-contact control training conditions. The error bars represent standard error of the mean. Please see Fig. S6 for a plot of this data with all data points visible.

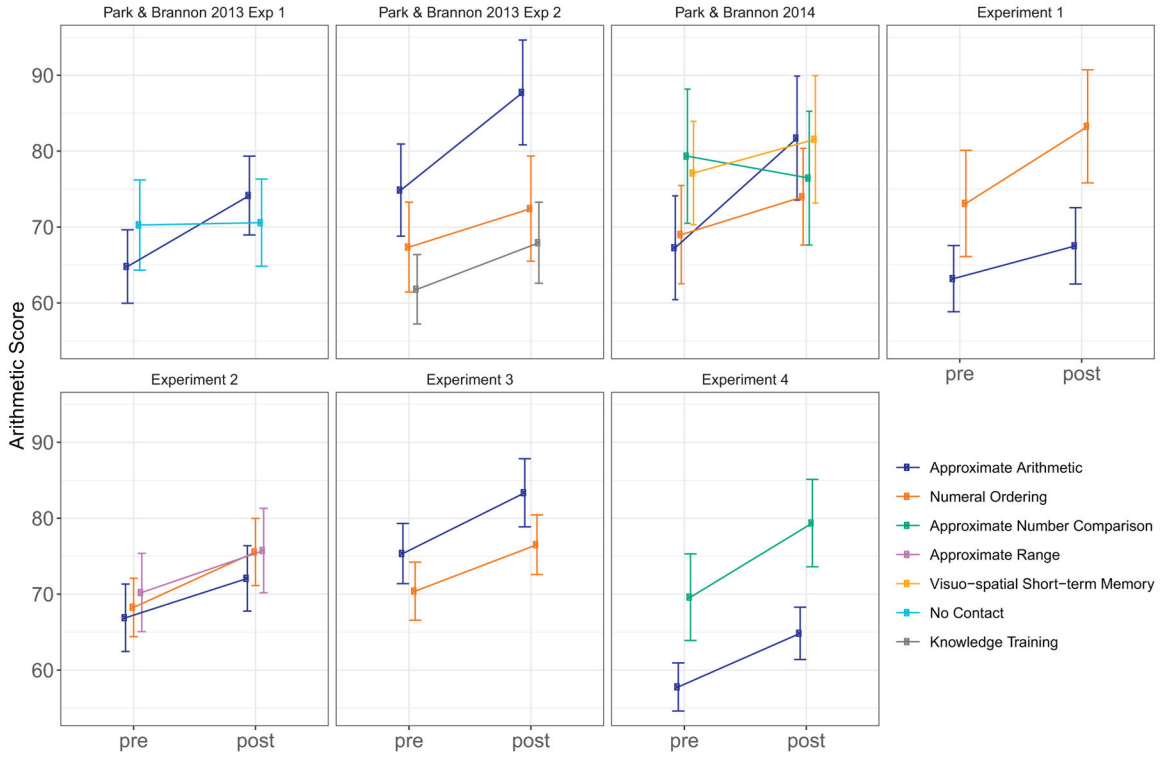

Fig. 3.

Pre and post test scores of the exact symbolic arithmetic test by experiment and condition. The outcome measure of this assessment is the number of arithmetic problems participants solved correctly in 10 min. Error bars indicate standard error of the mean. Points are jittered horizontally to maximize visibility. There were no significant differences in pretest score by condition. Experiment 4 includes one participant in the Approximate Number Comparison condition who scored 6 standard deviations above the mean (288 questions correct at pretest, 296 questions correct at post-test). This participant is included because their gain score is within the normal range. See also Supplementary Fig. S8 for a plot of pre and post test scores with all points visible.

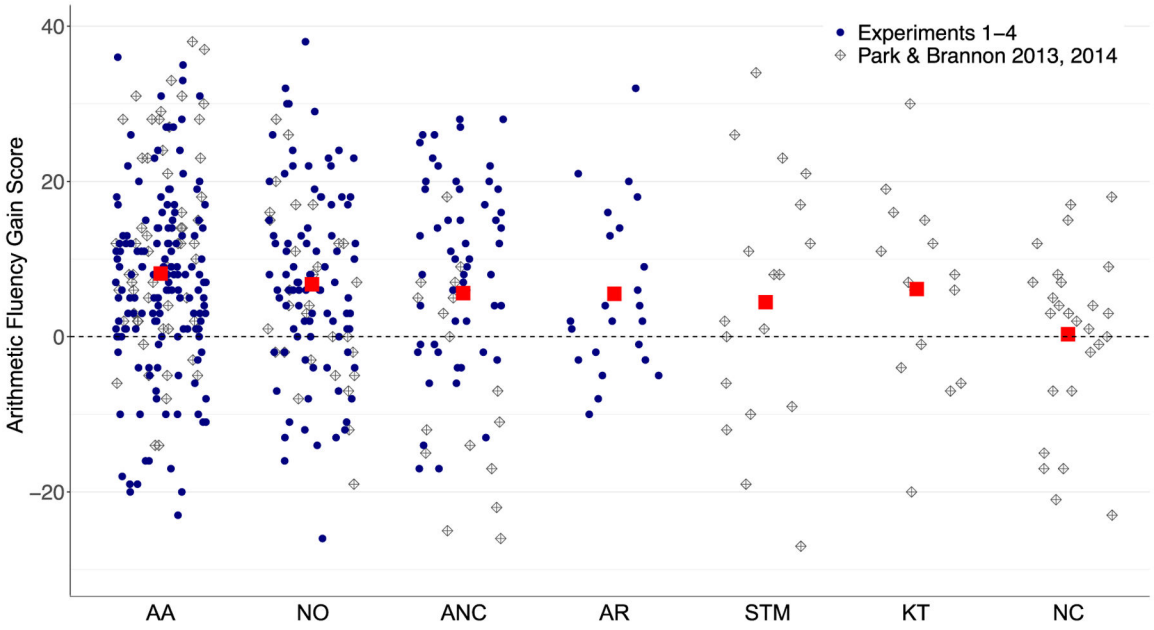

Fig. 4.

Arithmetic fluency gain score by condition collapsed across all experiments including the original Park and Brannon experiments. Each point reflects one participant's arithmetic fluency gain score. The points are randomly jittered horizontally to make all gain scores visible. The gain score is plotted in terms of the number of arithmetic questions solved correctly (post-test minus pretest). The dotted line indicates zero, which is the gain score of a participant who answered exactly the same number of arithmetic problems correct and pre and post-test. The red square indicates the mean gain score of participants in each training condition across all experiments. Data from participants in the original Park & Brannon, 2013, 2014 experiments are represented with diamonds. Data from participants in the current Experiments 1–4 are represented with dots. AA = approximate arithmetic, NO = numeral symbol ordering, ANC = approximate number comparison, AR = approximate range, STM = visuo-spatial short term memory, KT = knowledge training, NC = no contact.

Table 1

A methods summary of the four Experiments reported in the current paper, along with the experiments reported in Park & Brannon, 2013, 2014 for comparison. The days between pre and post test, age and gender of subjects, and the number of subjects are reported by condition. Each day of training consisted of a 25 min training session.

Dataset	Pre and post tests	Days of training	Training conditions	Mean days between pre and post test	Mean age of subjects (range)	Subject gender	Number of subjects
Park & Brannon, 2013	• Exact symbolic arithmetic • Vocabulary	10	Approximate arithmetic	10.9	22.4 (18.8–31.4)	9 M,17 F	26
Experiment 1			No contact control	11.3	22.9 (18.6–33.4)	6 M 20 F	26
Park & Brannon, 2013	• Exact symbolic arithmetic • Vocabulary	6	Approximate arithmetic	8.9	20.9 (18.7–23.8)	3 M 13 F	16
Experiment 2	• Numeral order judgement		Numerical symbol ordering	9.3	22.85 (18.8–31.9)	5 M 9 F	14
			Knowledge Training	9.1	22.17 (19–26.9)	6 M 10 F	15[a]
Park & Brannon, 2014	• Exact symbolic arithmetic • Vocabulary	6	Approximate arithmetic	9.2	21.4 (18.1–26.6)	7 M 11 F	18
Experiment 1	• Non-symbolic numerical comparison • Spatial 2-back test • Numeral order judgement		Numerical symbol ordering	9.1	21.4 (18.1–26.6)	6 M 11 F	17
			Approximate number comparison	9.2	21.9 (18.6–31.2)	7 M 11 F	18
			Visuo-spatial short term memory	9.3	22.6 (18.1–34.3)	8 M 9 F	18
Experiment 1	• Exact symbolic arithmetic • Numeral order judgement • Symbolic addition verification • Expectation matching • questionnaire (post-test only)	2	Approximate arithmetic	1.47	20.6 (18.1–24.9)	7 M 12 F	19
			Numerical symbol ordering	1.58	20.0 (18.6–22.4)	10 M 9 F	19
Experiment 2	• Exact symbolic arithmetic • Non-symbolic numerical comparison • addition verification (post-test only)	6	Approximate arithmetic	9.15	21.83 (18.8–27.8)	9 M 17 F 1 unreported	27
			Numerical symbol ordering	9.15	21.5 (18.5–34.1)	8 M 16 F 3 unreported	27

Dataset	Pre and post tests	Days of training	Training conditions	Mean days between pre and post test	Mean age of subjects (range)	Subject gender	Number of subjects
	• mental rotation • Numeral order judgement • rhyming test • Expectation matching • questionnaire (post-test only)		Approximate range	9.58	21.3 (18.7–30.4)	9 M 15 F	24
Experiment 3	• Exact symbolic arithmetic • vocabulary • Non-symbolic numerical comparison • Spatial 2-back test • Numeral order judgement • Expectation matching • questionnaire (post-test only)	6	Approximate arithmetic	10.65	23.09 (18.4–31.9)	13 M 34 F 1 unreported	48
			Numerical symbol ordering	10.33	24.63 (18.2–34.7)	19 M 24 F	43
Experiment 4	• Exact symbolic arithmetic • vocabulary • Non-symbolic numerical comparison • Spatial 2-back test • Numeral order judgement • Expectation matching questionnaire (post-test only)	6	Approximate arithmetic	10.22	22.4 (18.0–30.7)	16 M 37 F 2 unreported	55
			Approximate number comparison	10.18	22.6 (18.0–30.6)	13 M 36 F	56

[a] One subject was removed from the knowledge training condition of Park & Brannon, 2013 Experiment 2 due to the miscoding of this participant's training condition.