Title

Effects of non-symbolic arithmetic training on symbolic arithmetic and the approximate number system

Authors

Au, Jacky
Jaeggi, Susanne M
Buschkuehl, Martin

Publication Date

2018-04-01

DOI

10.1016/j.actpsy.2018.01.005

Peer reviewed

Published in final edited form as:
Acta Psychol (Amst). 2018 April ; 185: 1–12. doi:10.1016/j.actpsy.2018.01.005.

Effects of Non-Symbolic Arithmetic Training on Symbolic Arithmetic and the Approximate Number System

Jacky Au[*,1,2], **Susanne M. Jaeggi**[1,3], and **Martin Buschkuehl**[2]

[1]Department of Cognitive Sciences, University of California, Irvine, Irvine, CA, 92697, USA

[2]MIND Research Institute, Irvine, CA, 92617, USA

[3]School of Education, University of California, Irvine, Irvine, CA, 92697, USA

Abstract

The approximate number system (ANS) is an innate cognitive template that allows for the mental representation of approximate magnitude, and has been controversially linked to symbolic number knowledge and math ability. A series of recent studies found that an approximate arithmetic training (AAT) task that draws upon the ANS can improve math skills, which not only supports the existence of this link, but suggests it may be causal. However, no direct transfer effects to any measure of the ANS have yet been reported, calling into question the mechanisms by which math improvements may emerge. The present study investigated the effects of a 7-day AAT and successfully replicated previously reported transfer effects to math. Furthermore, our exploratory analyses provide preliminary evidence that certain ANS-related skills may also be susceptible to training. We conclude that AAT has reproducible effects on math performance, and provide avenues for future studies to further explore underlying mechanisms - specifically, the link between improvements in math and improvements in ANS skills.

Keywords

Approximate Number Sense; Cognitive Training; Arithmetic; Math; Symbolic; Number System; Numerical Distance Effect

1 Introduction

The Approximate Number System (ANS) is a primitive cognitive system present across many species, both human and non-human alike. It endows the individual with an intuitive, albeit approximate, understanding of magnitude, and underlies such common human faculties as estimating the number of apples on a tree or the number of jelly beans in a jar. This ability is apparent even in human infants prior to the onset of any formal numerical

[*]Corresponding Author: Jacky Au, 2201 Social & Behavioral Sciences Gateway Building, Department of Cognitive Sciences, University of California, Irvine, CA 92697, jwau@uci.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

instruction, and is thought to provide a natural template upon which to build an understanding of symbolic numbers (Lipton & Spelke, 2005; Mundy & Gilmore, 2009; Piazza, 2010).

Much behavioral evidence supports a close link between the representation of ANS numerosities and exact symbolic numbers, and suggests that the two share similar behavioral signatures. Most notably, both are susceptible to numerical distance effects such that identifying the larger of two quantities is more difficult the closer the quantities are together. For example, in a prototypical dot comparison task to measure ANS acuity, discriminating an array of 10 dots from an array of 12 is harder than discriminating 10 from 20 dots, and this distance effect can be observed in terms of both increasing reaction time as well as decreasing accuracy the smaller the ratio between the two dot arrays becomes (Dehaene, Dehaene-Lambertz, & Cohen, 1998). Similarly, a symbolic distance effect has robustly demonstrated longer reaction time latencies when identifying the larger of two closely spaced numbers such as 5 and 6, as opposed to relatively more distant numbers such as 5 and 9 (Moyer & Landauer, 1967). Additionally, controlling for the numerical distance between two quantities, a size effect also exists in that larger numbers or numerosities are more difficult (i.e., longer reaction times) to distinguish than smaller ones (Buckley & Gillman, 1974; Dehaene et al., 1998).

Taking this relationship one step further, the acuity of the ANS has also been shown to predict formal math ability (reviewed in Feigenson, Libertus, & Halberda, 2013). This relationship has been mainly explored in young children (Libertus, Feigenson, & Halberda, 2013a, 2013b; Odic et al., 2016), but exists throughout the school years (Halberda, Mazzocco, & Feigenson, 2008), and even correlates with SAT and GRE quantitative scores in adolescents and young adults (Dewind & Brannon, 2012; Libertus, Odic, & Halberda, 2012; Wang, Halberda, & Feigenson, 2017). Moreover, the link has been demonstrated in individuals with poor math ability (e.g., Mazzocco, Feigenson, & Halberda, 2011; Olsson, Ostergren, & Traff, 2016; Piazza et al., 2010), typical math ability (Feigenson et al., 2013), as well as precocious math ability (Wang et al., 2017), suggesting that the influence of the ANS on math is pervasive not only across a broad age range, but also across different levels of education and math proficiency. However, these findings are not without controversy, and several null reports have been published contesting the relationship between ANS and formal math, both in children as well as in adults (reviewed in Feigenson et al., 2013). The reasons for this inconsistency likely relate at least in part to psychometric differences across studies and low concurrent validity among ANS tests (Dietrich, Huber, & Nuerk, 2015; Gilmore, Attridge, & Inglis, 2011; Smets, Gebuis, Defever, & Reynvoet, 2014). Different tasks purporting to index the ANS often have low correlations with each other, and therefore, different studies may not always be measuring the same underlying construct. Nevertheless, throughout this noise, cumulative meta-analytic evidence still supports the existence of an overall small, but reliable correlation (r=.20 to .24) between math and ANS acuity (Chen & Li, 2014; Schneider et al., 2016), supporting the contention that the ANS is in fact related to mathematical and numerical knowledge.

The prospect of ANS plasticity is therefore of considerable interest, as it may implicate downstream effects on higher order skills. Although this effect is small, and certainly less predictive of later math performance than the more commonly studied symbolic processing of numbers (de Smedt, Noel, Gilmore, & Ansari, 2013), it still represents a heretofore largely untapped avenue for intervention. Moreover, intervention can occur at an unprecedentedly early age since the ANS is behaviorally present even in infancy (Starr, Libertus, & Brannon, 2013). From there, ANS acuity gradually increases throughout childhood (Halberda & Feigenson, 2008; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Piazza et al., 2010) and even throughout the young adult years, not peaking until around age 30 (Halberda et al., 2012), suggesting a high degree of plasticity. Additionally, education, particularly in quantitative fields, has also been shown to lead to a more refined ANS (Castronovo & Gobel, 2012; Halberda et al., 2008; Lindskog, Winman, & Juslin, 2014; Piazza, Pica, Izard, Spelke, & Dehaene, 2013). Therefore, ANS acuity, though innate, may also be highly susceptible to experience and environmental input. In fact, targeted interventions involving repeated practice on number sense tasks have sought to test this plasticity more specifically, demonstrating improved ANS acuity in typically developing children (Odic, Hock, & Halberda, 2014), improved acuity and number processing in dyscalculic children (Wilson, Revkin, Cohen, Cohen, & Dehaene, 2006), rapid learning effects in response to trial-by-trial feedback in healthy adults (Dewind & Brannon, 2012; Lindskog, Winman, & Juslin, 2013), and generalized magnitude discrimination improvements when coupling ANS exposure with transcranial random noise stimulation (Cappelletti et al., 2013).

Park and Brannon (2013, 2014) took this one step further and demonstrated that training to improve ANS skills via an approximate arithmetic training (AAT) task can also improve symbolic arithmetic skills among college students, as measured by addition and subtraction of Arabic numerals. Given the correlations between the ANS and mathematics performance throughout the school years, up to and including SAT and GRE scores (Dewind & Brannon, 2012; Libertus et al., 2012), this finding suggests a potential causal link between the ANS and mathematics that can be exploited by targeted training that fosters the bottom-up development of numeracy skills at a core, foundational level. This finding was later replicated among preschoolers with a standardized test battery of math achievement using a similar training intervention (Park, Bermudez, Roberts, & Brannon, 2016), and Wang et al. (2016) also concurrently found that even brief exposure to an ANS acuity task over a single session can improve formal math ability among preschoolers if the ANS trials are presented in a scaffolded manner (i.e., easier trials first). Despite these promising initial results, however, evidence for true plasticity at the level of the ANS has been criticized and is still inconclusive (Lindskog & Winman, 2016; Lindskog et al., 2013; Szucs & Myers, 2017), casting much uncertainty on what exactly mediates the improvements observed in math. One issue is that AAT, which involves the approximate addition and subtraction of dot clouds of varying numerosity, may be training additional processes beyond the ANS itself. Though Park & Brannon (2014) ruled out secondary processes such as visual working memory, covert symbolic arithmetic practice during the AAT, or general placebo effects, they were also not able to demonstrate any convincing training-related improvements on a measure of ANS acuity, and it is still an open question as to whether math improvements after AAT are specifically related to changes in the ANS *per se* (e.g., see Szucs & Myers, 2017).

The present study, therefore, has two goals. First, we attempt an independent replication of the transfer effects of AAT on symbolic arithmetic proficiency. Second, we seek to systematically explore direct transfer effects of AAT to ANS-related skills. With respect to the second goal, we aim to improve on the methodology used by Park & Brannon in several ways. First, while Park & Brannon (2014) used a single measure (a nonsymbolic comparison task) to index the ANS, we use a battery of different tasks, evaluating both nonsymbolic and symbolic tests of comparison, estimation, and nonverbal counting. Evidence suggests that the ANS may not represent a unitary construct, and that different metrics do not correlate well with each other (Gilmore et al., 2011; Smets et al., 2014). Therefore, a valid assessment of training-related ANS change would likely require multiple measures. Moreover, our use of both nonsymbolic as well as symbolic versions of each of our tasks allows an evaluation of both specific and general transfer to number sense. If improvements in math are truly a result of specific improvements in the ANS, then these improvements in nonsymbolic discrimination must also be generalizable to the symbolic domain as well. Finally, we seek to maximize the chances of transfer by creating outcome measures that more closely mirror the relevant characteristics of the training regimen. For example, one issue with the nonsymbolic comparison task used in Park & Brannon (2014) to assess near transfer to the ANS is that it involved the comparison of ratios typically much smaller than what was trained. Fig. 2 of Park & Brannon (2014) shows a log difference level of just over 0.5 at the end of six training sessions, which corresponds to discriminating dot arrays that are separated by approximately a ratio of 1.5 to 1. However, their nonsymbolic comparison task tested participants on ratios that were almost all below 1.25 to 1, a range on which they received very little training. Therefore, the tasks used in the present study, including our version of the nonsymbolic comparison task, incorporate magnitude information designed to contain greater overlap with the trained numerosities, and our statistical analyses are designed to investigate the degree to which this matters by systematically evaluating group differences across different magnitude ranges.

Another issue is that Park & Brannon controlled for continuous perceptual cues such as average dot size and total surface area in their transfer task, but not in the training task. Though such non-numerical stimulus control has recently become common practice in the literature (c.f., Dietrich et al., 2015), and is arguably a more pure way to measure the abstraction of numerical information, unconfounded by other continuous perceptual cues, this makes the task much harder for participants (Agrillo, Piffer, & Bisazza, 2011; Dietrich et al., 2015; Gebuis & Reynvoet, 2012b), and may not entirely engage the same cognitive processes that were trained considering that the training task did not control for such perceptual cues. In order to maximize chances of detecting transfer effects, it is important to increase process overlap with the trained task (c.f., Jaeggi et al., 2010; Loosli, Buschkuehl, Perrig, & Jaeggi, 2012; Lustig, Shah, Seidler, & Reuter-Lorenz, 2009). Therefore, this required making a design choice on our part to either control for non-numerical cues in our training task, or to keep the training task as is and remove such controls from the transfer tasks. We opted for the latter choice in order to keep the training as consistent as possible to that of Park and Brannon (2013, 2014), reasoning that this approach would maximize chances of replicating the transfer effects to symbolic arithmetic proficiency, as any attempt to evaluate the underlying mechanisms of training would otherwise be be moot.

Moreover, it has been demonstrated that both humans and non-human species such as fish learn faster and are more accurate with approximate discrimination when they can merge redundant information from several sources such as numerosity and various perceptual cues (Agrillo et al., 2011; Gebuis & Reynvoet, 2012b). It has further been suggested that a true approximate number sense may not exist in humans independent of these perceptual cues, which in naturalistic environments, are virtually always confounded with numerosity (Gebuis & Reynvoet, 2012a). Controlling for such perceptual cues, therefore, may be fruitful in basic research aimed at investigating the underlying components of the ANS, but may arguably not be best suited for training applications seeking to maximize the efficacy of approximate discrimination ability. Although this imposes interpretive limitations on any transfer effects found with our nonsymbolic ANS outcomes, the present study therefore does not control for perceptual variables in order to maintain ecological validity and maximize process overlap with the training regimen.

2 Methods and Materials

2.1 Participants

Participants between the ages of 18 and 35 were eligible to participate and were recruited from the University of California, Irvine undergraduate and alumni community. Additionally, participants were required to have received their primary schooling in English in order to control for potential language differences in numerical cognition. Sixty-seven individuals were enrolled, seven withdrew after the pre-test, and three were excluded as outliers based on their training data (see below). In the end, 57 participants were included in the final sample (mean age ± SD: 21.08 ± 1.8 years, range = 18–26 years). All research procedures were approved by the Institutional Review Board and each subject signed an informed consent.

2.2 General Procedure

We used a between-subjects pre-test-post-test intervention design and randomized participants into one of two intervention groups after pre-test (Fig. 1). Twenty-seven received AAT and 30 received training on a control task that required answering vocabulary and general knowledge questions. Both groups trained at home on their respective tasks using software installed on their personal laptops, and both groups were told vaguely that the training was designed to improve general cognition on a foundational level. The intervention period consisted of 7 consecutive daily at-home sessions (including weekends and holidays). Participants were expected to email their training data upon the completion of each session so that progress could be monitored. Reminder emails were sent after each missed day, and skipping more than one day in a row resulted in exclusion from the study. All participants returned to the laboratory for post-test the day after their last training session. All laboratory tasks were conducted on a Dell PC desktop on a 19 inch monitor with a resolution of 1280×1024. Participants were compensated with either $40 or course credit. Additionally, all participants were entered into a lottery for $100 upon conclusion of data collection. Extra lottery tickets were earned based on training performance.

2.3 Training Tasks

2.3.1 Approximate Arithmetic Training—The training task used was programmed in PsychoPy (Peirce, 2009) and was based on the specifications laid out in Park & Brannon (2013). On each trial, participants were asked to approximately add or subtract two dot arrays (see Fig. 2 for schematic and description), which were presented each for only 1,000ms in order to prevent counting. Feedback was given after each trial, and difficulty was adjusted after each block by varying the numerical distance (ratio of dots between the larger array and the smaller one) either between answer choices (Fig. 2C) or between the correct answer and the reference array (Fig. 2D). Difficulty was adjusted separately for each answer format, and was calculated on a log-base 2 scale, with an initial value of 1.5 at the beginning of training (i.e., dots separated by a ratio of 2.83 to 1). This value was increased by .1 if accuracy on a block dropped below 70% and was decreased by .15 if accuracy was greater than 85%. Each block consisted of 20 trials, and a single training session consisted of 10 blocks. Participants were instructed to respond as accurately as possible, with no specific instructions regarding speed. However, the answer choices disappeared after one second, so some baseline level of speed was encouraged. Nevertheless, the participant was still able to input his or her response even after the answer choices disappeared and the next trial did not begin until after a response was made. The primary dependent variable for analysis was the average log difference level achieved per session. Although this log difference adaptivity scale was adapted from Park & Brannon (2013), and all analyses in the current report are based on this scale, qualitative interpretation throughout the rest of the manuscript will be based on their ratio conversion, which is a more intuitive metric.

2.3.2 Active Control Training—The active control group trained on a general knowledge task based on the task used in Jaeggi et al. (2014) and similar to what Park & Brannon (2013) also used. The task presented GRE-type general knowledge, vocabulary, and trivia questions. Each question was presented in the center of the screen, along with four answer alternatives. Feedback was provided after the participants' response, and incorrect responses were followed by the correct answer along occasionally with additional facts related to the question. Questions answered incorrectly were presented again in the beginning of the next session in order to evoke a learning experience. Also with this task, the emphasis was on accuracy, not speed, although a generous time limit of 15 seconds was afforded to participants to make a decision before a trial was automatically marked as incorrect.

2.4 Outcome Measures

With the exception of the arithmetic task, all outcome measures consisted of both a symbolic and nonsymbolic version. Order of outcomes was fixed for all participants and consisted of: nonsymbolic comparison, symbolic arithmetic, symbolic estimation, nonsymbolic estimation, nonverbal counting of interleaved symbolic and nonsymbolic magnitudes, and finally symbolic comparison. Each task was preceded by instructions and practice, and participants were instructed to complete all tasks as quickly but as accurately as they could.

2.4.1 Symbolic Arithmetic—The symbolic arithmetic task was modeled in part after the design used in Park & Brannon (2013, 2014), and as used previously in our lab (Mohammed et al., in press). Trials consisted of addition or subtraction of two or three operands, ranging between 11 to 244. Correct answers ranged from 11 to 284. Stimuli were presented in Arial font, with a font size of 75 pixels. Prior to the task, participants were exposed to four sample problems of representative difficulty and were not allowed to continue until they demonstrated understanding of how to carry out the operations. Instructions and guidance were given as needed. They then were given practice typing in 20 random numbers displayed on the screen, in order to familiarize their fingers to use the number pad. The task consisted of 80 trials in total, with a brief break given halfway through. Although participants were generally instructed to be quick and accurate, they were under no specific time pressure or time limit when answering each question. Unbeknownst to participants, the task timed out after 25 minutes in order to prevent excessive fatigue for slow math performers. This time limit was generous and 87% of participants finished the task within the limit. Each trial was randomly generated, but the task was balanced to contain 10 three-operand subtraction trials with a borrow operation, 10 three-operand addition trials with a carry operation, 20 two-operand subtraction trials with a borrow operation, 20 two-operand addition trials with a carry operation, 10 three-operand subtraction trials without a borrow operation, and 10 three-operand addition trials without a carry operation. The primary dependent variables were percent accuracy and reaction time, as measured by the time between stimulus onset and when participants pressed “enter” to submit their answer.

2.4.2 Comparison—The nonsymbolic comparison task was designed after standard numerosity comparison tasks and simultaneously presented two arrays of dots for 1,000ms, each ranging in numerosity from 9 to 30. Participants were instructed to indicate which array contained more dots as quickly as possible via a keyboard press (“A” or “L”). Dots were presented in white color against a black background, and dot sizes varied randomly from 4 to 8 pixels in diameter from trial to trial, but were constant for all dots within a trial. Each presentation was preceded by a central fixation cross for a jittered duration between 500 to 1,000ms. The ratios used between numerosities were 1.5:1, 1.25:1, and 1.2:1. For simplicity, further reference to ratios throughout this manuscript will drop the right-hand denominator in the notation (i.e., 1.5, 1.25, and 1.2). These ratios were chosen in order to be inclusive of the presumed training range, as informed by Park & Brannon (2013, 2014) where participants’ training hovered around an average ratio of 1.5 for most sessions (Fig. 3). We expected stronger transfer effects to occur on ratios that received more training. Participants underwent eight blocks of 16 trials each for a total of 128 trials. The dependent variable was percent accuracy, which is a more intuitive metric than the conventional weber fractions used in the literature, and provides almost identical information (Szucs, Nobes, Devine, Gabriel, & Gebuis, 2013).

The symbolic comparison task (Moyer & Landauer, 1967) sequentially presented a series of Arabic digits between 4 and 42, a range of numerical values inclusive of the 9 to 36 range of dot numerosities used during training. Stimuli were presented in Arial font with a font size of 75 pixels. Participants were instructed to compare each number to a reference number, 23, and indicate by keypress (“A” or “L”) as quickly as possible whether the presented number was less than or greater than the reference. Although the range of stimuli is broad, we expected the strongest distance effects to occur with the numbers closest to the reference (e.g., Hinrichs, Yurko, & Hu, 1981). Accordingly, we also hypothesized that if AAT is able to reduce distance effects, it would manifest most noticeably on the numbers nearest to the reference. A fixation cross was presented prior to each number for a jittered duration between 500 to 1,000ms. Numbers were grouped together for analysis according to their numerical distance from the reference. For example, both 22 and 24 were grouped together because they are both a distance of one away from the reference. Numerical distances of two through four were similarly grouped, and numerical distances beyond that were grouped in clusters of five (i.e., 5–9, 10–14, and 15–19). Numbers within a distance of four from the reference were each presented five times throughout the task. All other numbers were presented three times, creating a total of 130 total trials. The dependent variable was the median reaction time of all correct responses in each number group.

A second set of numbers was presented after the first 130 whole number trials, which consisted of 29 fractions compared to the reference 3/5. The fractions were evenly centered around the reference, and were taken from Schneider and Siegler (2010). No attempt was made to match the magnitude of these stimuli with the training task due to the inherent incompatibility of fractional values. Another practice round preceded these trials and the dependent variable was accuracy rather than reaction time as used in the symbolic comparison task described above since the increased difficulty of fractional magnitude estimation afforded sufficient variability in accuracy scores.

2.4.3 Estimation—The estimation tasks measured the degree of mapping between mental representations of nonsymbolic and symbolic quantities (i.e., mapping between approximate and exact number systems). The nonsymbolic variant of this task required the estimation of nonsymbolic quantities by generating symbolic numbers (i.e., an array of dots was presented on the screen and participants were asked to estimate the number of dots by inputting a number on the keyboard). The number of dots displayed varied from trial to trial and comprised the following target numerosities, binned into three groups based on size – small: 13, 14, 16, 18, medium: 21, 23, 26, 29, and large: 44, 48, 53, 57. The small and medium targets purposefully overlapped with numerosities in the training range, while the large targets contained numerosities that were not specifically trained, with the hypothesis that stronger transfer effects would occur in the small and medium bins. Dot sizes were homogenous within an array (ranging in diameter from 4 to 8 pixels and presented in white color against a black background), but varied randomly from trial to trial. Each trial was preceded by a 750ms central fixation cross. There were 4 blocks of 24 trials each. Each target numerosity was presented twice per block. The dependent variable was the absolute value of the error rate, calculated as a percentage deviation from the target $\left(\frac{R-T}{T}\right)$, where R is the median response per target, and T is the target.

The symbolic variant followed the same parameters described above except the encoding/response process was reversed. That is, participants were asked to estimate the magnitude of a symbolic number, presented in Arial font with a font size of 75 pixels, by creating a nonsymbolic representation (i.e., a dot array). These dot arrays were generated by using the scroll button on a standard computer mouse. Each scroll generated between 1–25 dots in a semi-random manner that correlated with the intensity of the scroll such that faster scrolls generated more dots. This was done by measuring the number of clicks (c) per scroll, and generating a random integer that lay between $c^{1.5}$ and $c^{1.8}$. The maximum number of clicks per scroll was capped at 6. This random algorithm was adopted in order to discourage participants from using any counting strategies while generating dot arrays. Dots could be either added or removed from the screen in this manner. Participants submitted their final answer with a click on the left mouse button when they felt they had generated an appropriate number of dots.

2.4.4 Nonverbal Counting—The nonverbal counting task is an analogue measure of ANS acuity commonly employed in the animal literature. The present iteration of the task was partially adapted from Whalen et al. (1999) and Cordes et al. (2001) and involved presenting either a symbolic (Arabic numeral) or nonsymbolic (dot array) number/ numerosity to participants and asking them to count nonverbally by pressing the spacebar a corresponding number of times as quickly as possible. They were simultaneously required to perform an articulatory suppression task by repeating the word “California” out loud in order to discourage subvocal counting. An experimenter was present to enforce compliance. Arabic numerals were presented in white Arial font, with a font size of 75 pixels, while dot arrays consisted of white dots varying in diameter between 4 to 8 pixels across trials. All stimuli were presented against a black background. Symbolic and nonsymbolic trials were interleaved together in alternating fashion, and participants were informed what trial type they were receiving before each trial, and proceeded by pressing a key whenever they were ready. In both trial types, the stimulus appeared on screen for only 1,000ms, but participants were able to continue inputting their answers until they felt they had achieved the requisite number of keypresses. They pressed “Enter” to submit their answer when done. Target numbers in the symbolic trials were 11, 14, and 23. Target numerosities in the nonsymbolic trials were 13, 17, and 26 dots. All numbers or numerosities overlapped with trained magnitudes, as time constraints made it impractical to use stimuli above the training range. We therefore had no specific hypothesis about differential transfer effects for this task. Each target was presented once per block over ten blocks, with 60 trials total. The dependent variable was the absolute value of the error rate, as used in the estimation tasks.

2.5 Analysis Plan

Statistical analyses were conducted using STATA version 13 (StataCorp, 2013). Where applicable, data were analyzed using the median of variables of interest rather than the mean in order to protect against the undue influence of large outliers. The results section is organized as follows. First, we screened for individuals who presented as outliers in the training data, excluding participants who exhibited signs of non-compliance with their particular training regimen (see Results). This was especially pertinent given that participants trained at home in an unsupervised environment, and allowed us to reduce the amount of noise in the data. Next, in order to analyze whether a common construct, such as ANS performance, underlies performance on our tasks, we ran correlations between all baseline measures, including session one of the AAT. We then moved on to the analysis of the training data using a repeated measures analysis of variance (ANOVA) to investigate task-related improvements over the seven training sessions.

The main analyses, however, consisted of a systematic evaluation of each outcome measure using both confirmatory and exploratory analyses. We started with a confirmatory analysis of the symbolic arithmetic task, in order to replicate previous effects (Park & Brannon, 2013, 2014), and then moved on to exploratory analyses of the comparison, estimation, and counting tasks. Before assessing transfer for each task, we first sought to validate various psychometric properties of each test in two ways. First we calculated Cronbach's alpha (e.g., Bland & Altman, 1997) to measure the internal consistency of each test by dividing sequential test items evenly into three bins and calculating the average inter-item reliability of all pairwise combinations of bins. Secondly, where appropriate, we sought to verify the existence of theoretically-grounded classical signatures of the ANS in each task. Such signatures include distance or magnitude effects (Buckley & Gillman, 1974; Dehaene et al., 1998; Moyer & Landauer, 1967) as well as stable coefficients of variation in estimation and nonverbal counting tasks (Castronovo & Gobel, 2012; Cordes et al., 2001; Whalen et al., 1999). A stable coefficient of variation indicates scalar variability wherein the variability in mental magnitude representation increases proportionally with the target magnitude such that their ratio remains constant over different target magnitudes. If these properties were confirmed, we went on to assess transfer with analyses of covariance (ANCOVA) comparing post-test performance between groups, using pre-test performance as a covariate (Dugard & Todman, 1995; Huck & Mclean, 1975). In the case of either a significant group effect, or a significant interaction, we reran additional ANCOVA analyses at each target magnitude individually. Although these are fairly liberal criteria for running follow-up analyses, we decided this was important due to the exploratory nature of our ANS-related tasks. For similar reasons, we did not correct for multiple comparisons (c.f., Simons et al., 2016), but caution against over-interpretation of the exploratory aspects of our results. As a complementary analysis to the ANCOVAs, we also calculated change-from-baseline effect sizes for both groups, accounting for the correlation (r) between pre and post-test, using the formula: $\frac{(\mathrm{Mean}_{\mathrm{Post}} - \mathrm{Mean}_{\mathrm{Pre}})}{\sqrt{\mathrm{SD}^{2}{}_{\mathrm{Pre}} + \mathrm{SD}^{2}{}_{\mathrm{Post}} - 2r * \mathrm{SD}_{\mathrm{Pre}} * \mathrm{SD}_{\mathrm{Post}})}}$. Finally, regression analyses were conducted seeking to explain each significant transfer result (i.e., the outcome variables) as a function of AAT performance (i.e., the predictor variable), controlling for baseline performance.

3 Results

3.1 Outlier Analysis and Evaluation of Baseline Differences

Outliers in the training data were identified by examining the fit (R^2) of each individual training curve (see Training Gains below for training curve analysis) over all seven sessions to the average training curve, separately for both ANS and control groups. A poor fit was defined as an R^2 value that was more than 2 median absolute deviations away from the overall median (Leys, Ley, Klein, Bernard, & Licata, 2013). This resulted in the identification of three low-performing outliers in the AAT group, all of whom performed actually worse at the end of training than in the beginning. Given the easy level at which this training regimen starts, a level considerably below the ability level of typically developing adults (Lindskog & Winman, 2016), and given the fact that our sample consisted of mainly well-educated college students, we concluded that these individuals were not performing the task properly during the training period. Therefore, these individuals were excluded from all analyses. The same procedure resulted in the identification of two high-performing outliers in the control group; however, they were retained in the sample in order to reduce the risk of bias. This resulted in a total of 27 AAT participants and 30 control participants in the final sample. At post-test, one data point from the AAT group was lost on the Symbolic Arithmetic and Nonsymbolic Comparison tasks due to technical errors, and one participant failed to complete the Symbolic Comparison and Nonverbal Counting tasks due to time constraints. These participants' data were also removed from pre-test, resulting in n=26 for each of those tasks. One-way ANOVAs revealed no statistically significant differences in pre-test performance on any measure between the AAT group and the control group (p's < 0.21). Detailed pre-test (and post-test) data, as well as reliability estimates and effect sizes, for each task are included in the supplementary online materials.

3.2 Correlations between Pre-Test and Approximate Arithmetic Training

Table 1 shows a correlation matrix of all pre-test data and performance in the first training session. No systematic pattern of correlations emerged, but a few isolated tasks showed significant correlations: Nonsymbolic Comparison vs. Fraction Comparison, $r(55)=.31$, $p=.02$; Symbolic Nonverbal Counting vs. Nonsymbolic Nonverbal Counting, $r(55)=.57$, p<.01; Fraction Comparison vs. Training Session 1, $r(25)=.43$, $p=.02$; Nonsymbolic Estimation vs. Training Session 1, $r(25)=.50$, $p<.01$.

3.3 Training Gains

Participants significantly improved on the trained task, as confirmed by a repeated measures ANOVA on average log difference levels, which revealed a significant main effect of session, $F(6, 156) = 49.48$, p<.01, $\eta_p^2 = 0.66$ such that participants achieved lower (i.e., more difficult) ratios as sessions progressed. On average, from the first to the last training day, the average ratio of dots that participants successfully resolved improved from 2.26 to 1.56. Our results are superimposed on those from the original Park & Brannon (2013) study (see Fig. 3). For ease of comparison, axes are labeled both in terms of ratio as well as log difference levels.

3.4 Transfer to Symbolic Arithmetic

Before assessing transfer, we measured the inter-item reliability of the Symbolic Arithmetic task. Cronbach's alpha was 0.76 at pre-test and .80 at post-test. To assess transfer of training gains, we compared mean accuracy scores on symbolic arithmetic between the ANS and control groups. On average, participants in the AAT group achieved 84% accuracy at pre-test (SD: 6%) and 88% at post-test (SD: 7%). This translates to solving 67.44 problems correctly at pre-test and 70.46 problems at post-test. In contrast, the control group showed no improvement, with 86% accuracy at pre-test (SD: 10%) and 86% at post-test (SD: 9%), solving an average of 68.68 problems at pre-test and 68.75 problems at post-test. An ANCOVA controlling for pre-test scores revealed significant differences in adjusted post-test mean accuracy scores, with the AAT group outperforming the controls, $F(1, 53) = 4.30$, $p = .02$, $\eta_p^2 = 0.08$ (one-tailed). A similar analysis with reaction time as the dependent variable revealed no group differences, $F(1, 53) = .60$, $p=.44$, $\eta_p^2 = 0.01$). Effect size calculations revealed an accuracy gain of d=0.62 in the AAT group, compared to d=0.01 in the control group. Our results are juxtaposed with those of Park & Brannon (2013) in Fig. 4.

3.5 Transfer to ANS-related Outcomes

3.5.1 Comparison—Cronbach's alpha for the Nonsymbolic Comparison task was 0.59 at pre-test and 0.65 at post-test. Near transfer of training gains onto Nonsymbolic Comparison was evaluated with a 2×3 ANCOVA on post-test accuracy, using the between-subjects factor Condition (AAT, control), and the within-subjects factor Ratio (1.5, 1.25, 1.2), and pre-test scores as covariates. As expected, we replicated classic distance effects by finding a main effect of Ratio: $F(2, 107) = 42.47$, $p<.001$, $\eta_p^2 = .44$, which exhibited a significant linear contrast, $F(1, 107) = 78.39$, $p<.001$, such that greater accuracy was found in higher ratios (Fig. 5a). Importantly, however, we also found a main effect of Condition, $F(1, 54) = 5.76$, $p=.02$, $\eta_p^2 = .12$, with the AAT group achieving greater overall accuracy than the control group, but no Condition×Ratio interaction, $F(2, 107) = 1.47$, $p = .23$, $\eta_p^2 = .03$. In order to test our hypothesis of ratio-specific training and transfer effects, we ran individual ANCOVAs for each ratio separately. Significant differences in favor of the AAT group were observed for the 1.25 and 1.5 ratios, $F(1, 53) = 4.45$, $p = .04$ and $F(1, 53) = 9.04$, $p < .01$, respectively, but not for the 1.2 ratio, $F(1, 53) = .27$, $p = .61$. Finally, as a complementary measure, we calculated change-from-baseline effect sizes (Cohen's d) for each group at each ratio (Fig. 5b). This analysis revealed that the ANCOVA effect was *not* driven by improvements in the AAT group, because no significant improvements were observed at any ratio (all p's>0.39; all d's < .18), but the control group exhibited a significant decrement in performance at the 1.5 ratio, d=−0.46, p=.02.

Fig. 6 shows time on task during the seven sessions of AAT as a function of ratio, which was binned into deciles between 1 and 3, with each bin spanning a range of 0.2. The figure demonstrates that the majority of training time was spent on trials with ratios around 1.3 to 1.5, roughly corresponding with the range of ratios that showed the strongest post-test differences between groups on the Nonsymbolic Comparison task. In contrast, there is a sharp drop-off below that range, with relatively little training time spent on ratios of 1.2 or below, corresponding with a complete lack of improvement on the 1.2 ratio on the Nonsymbolic Comparison task. Unfortunately, due to the strong dependencies in the time on task data (i.e., there exists a complex trade-off in time on task between a given ratio and its neighbors due to the nature of the adaptive algorithm), more sophisticated regression analyses seeking to explain variance in the Nonsymbolic Comparison task based on training time on a particular ratio are not very informative. However, see Regression Analyses section further below for additional analyses.

For the Symbolic Comparison task, Cronbach's alpha was calculated to be .96 at pre-test and .97 at post-test. Transfer was assessed with a 2×7 ANCOVA on post-test reaction times, using the between-subjects factor Condition (AAT, control), and the within-subjects factor Distance (1, 2, 3, 4, 5–9, 10–14, 15–19), using pre-test scores as covariates. Again, we replicated classic distance effects by observing a main effect of Distance, $F(6, 317) = 18.58$, $p < .001$, $\eta_p^2 = .26$, with a significant linear contrast, $F(1, 317) = 102.28$, $p<.001$ such that smaller numerical distances exhibited greater reaction time latencies (Fig. 7). Transfer was demonstrated with a main effect of Condition, $F(1, 53) = 6.02$, $p = .02$, $\eta_p^2 = 0.12$ in favor of faster reaction times in the AAT group. There was no Condition×Distance interaction, $F(6, 317) = 1.72$, $p = .12$, $\eta_p^2 = .03$. Again, additional individual ANCOVAs for each distance were calculated to test whether transfer effects would be strongest at numerical distances closest to the reference. Our results supported our hypothesis, showing that the effects were largely driven by the numerical distance of 1 (i.e., the nearest neighbors to the reference, 22 and 24), $F(1, 52) = 6.17$, $p = .02$, $\eta_p^2 = .11$. None of the other distances (2 through 19) reached significance (all p's > .13). Change-from-baseline effect sizes were also calculated for both groups at each numerical distance. In line with the ANCOVA analyses, the AAT group improved significantly more than the control group at a numerical distance of 1 (d=0.63 vs. d=.02, p=.02). No other significant comparisons were observed (all p's > .17).

No significant effects were found on accuracy rates of Fraction Comparison: ANCOVA condition effect, $F(1, 54) = .18$, $p = .67$, $\eta_p^2 < .01$). Cronbach's alpha for Fraction Comparison was .71 at pre-test and .67 at post-test.

3.5.2 Estimation—Cronbach's alpha for the Nonsymbolic Estimation task was .76 at pre-test and .91 at post-test. We calculated the coefficients of variation of participants' responses, expecting them to remain constant across target sizes (small, medium, and large). These coefficients are traditionally calculated by dividing the standard deviation of all responses by the mean of all responses for each target numerosity (or number, for the symbolic task). To protect against the undue influence of outliers in the data, we modified this equation by calculating the median absolute deviation divided by the median. We then calculated a repeated measures one-way ANOVA on coefficients of variation with the factor Target Size (small, medium, large) for pre-test only. As expected, there was no main effect of Target Size, $F(2, 112) = .74$, $p=.48$, $\eta_p^2 < .01$, indicating stable coefficients of variation.

We then proceeded to analyze transfer by calculating error rates for each participant and subjecting the data to a 2×3 ANCOVA with the between-subjects factor Condition (AAT, control) and the within-subjects factor Target Size (small, medium, large), controlling for pre-test performance. As expected, there was a main effect of Target Size, $F(2, 109) = 5.98$, $p < .01$, $\eta_p^2 = .10$, indicating greater error with increasing size with a significant linear contrast, $F(1,109) = 11.27$, $p<.01$. There was also a significant interaction between Condition and Target Size, $F(2, 109) = 3.66$, $p = .03$, $\eta_p^2 = .06$, where the AAT group has higher error rates at low target sizes, but lower error rates at high target sizes. However, there were no main effects of Condition, $F(1, 55) = .19$, $p = .66$, $\eta_p^2 = .01$ and individual ANCOVAs at each target size similarly returned null results (all p's > .14).

For the Symbolic Estimation task, Cronbach's alpha was .99 at both pre-test and post-test. Analyses of coefficients of variation revealed a main effect of Target Size, $F(2, 112) = 11.75$, $p < .01$, $\eta_p^2 = .17$, suggesting inconsistency across target sizes. Follow-up pair-wise t-tests revealed all size bins to be different from each other (small vs. medium: $t(112) = -2.33$, $p = .02$; medium vs. large: $t(112) = -2.15$, $p=.03$; small vs. large: $t(112) = -4.07$, $p<.01$). Due to the unstable nature of the coefficients of variation, we deemed the validity of the symbolic estimation task to be questionable at best, and therefore no further analyses were run.

3.5.3 Nonverbal Counting

For the Nonsymbolic Nonverbal Counting task, Cronbach's alpha was calculated as .93 at pre-test and .96 at post-test. Again, coefficients of variation were calculated with a one-way repeated measures ANOVA on the factor Target (13, 17, 26), which revealed a marginal effect ($F(2, 110) = 2.81$, $p = .06$, $\eta_p^2 = .05$). Follow-up tests revealed this to be driven by a difference between the smallest target, 13, compared to the others, $t(110) = 2.33$, $p = .02$. Since the omnibus effect is only marginal, we still proceeded with transfer analyses, though we ran our model with and without the problematic target.

Transfer effects were analyzed by subjecting error rates to a 2×3 ANCOVA with the between-subjects factor Condition (AAT, control) and the within-subjects factor Target, controlling for pre-test performance. There was a main effect of Target, $F(2,107) = 5.05$, $p<.01$, $\eta_p^2 = .08$, with an unexpected quadratic contrast, $F(1,105)=5.31$, $p<.01$, in which the middle target, 17, has a lower error rate than the small and large targets. There was no main effect of Condition, $F(1,54) = 1.58$, $p = .21$, $\eta_p^2 = .07$, and no Condition×Target interaction, $F(2, 107) = 1.12$, $p=.33$, $\eta_p^2 = .02$. Due to the marginal effect of Target on the coefficients of variation, as noted above, we reran this model after removing the smallest target. This restored the expected linear trend in the Target factor, $F(1, 52) = 10.84$, $p<.01$, with greater error rate in the high target compared to the middle target. However, the qualitative results did not change, Condition: $F(1,54) = 1.33$, $p = .25$, $\eta_p^2 = .12$, Condition × Target: $F(2, 107) = 3.41$, $p=.07$, $\eta_p^2 = .06$.

For the Symbolic Nonverbal Counting task, Cronbach's alpha was calculated as .95 at pre-test and .96 at post-test. Coefficients of variation were analyzed with a one-way repeated measures ANOVA on the factor, Target (11, 14, 23), which revealed no effects, $F(2, 100) = .66$, $p = .52$, $\eta_p^2 = .01$, indicating stable coefficients of variation across target magnitudes. Transfer analyses were run the same way as described above in the nonsymbolic task. Contrary to expectations, there was no main effect of Target, $F(2, 107) = 2.66$, $p=.08$, $\eta_p^2 = .05$, suggesting equal error rates irrespective of target size. We also found no main effect of Condition, $F(1,54) = .31$, $p=.58$, $\eta_p^2 =.01$, and no Condition×Target interaction, $F(2, 107) = .96$, $p = .39$, $\eta_p^2 = .02$.

A detailed summary of all means, standard deviations, and effect sizes for all transfer measures are available in the supplementary materials (Tables S1 – S3).

3.6 Regression Analyses

Next, we sought to establish whether a relationship existed between training performance on the approximate arithmetic task and transfer. To do so, we ran a separate regression model for each significant Condition main effect described above (Symbolic Arithmetic, Symbolic Comparison, Nonsymbolic Comparison), using training performance as the predictor variable, post-test performance as the outcome, and controlling for pre-test performance on the particular transfer outcome. Training performance was indexed by taking the average log difference level across all seven sessions. We found that training performance was only a significant predictor for the Nonsymbolic Comparison task, $\beta = 0.55$, $t(25) = -3.21$, $p < .01$, but not Symbolic Comparison, $\beta = 0.04$, $t(25) = 0.3$, $p=.77$ nor Symbolic Arithmetic, $\beta = 0.07$, $t(25) = 0.41$, $p=.68$. This relationship explained 29% of the variance in post-test performance of the Nonsymbolic Comparison task, $F(2, 23) = 5.49$, $R^2=.29$. Further details are provided in Table 2. As an additional control analysis (not shown in Table 2), we also regressed pre-test Nonsymbolic Comparison scores on training performance, and confirmed that no predictive relationship existed prior to intervention, β=.08, $t(25) = 0.4$, $p = .69$, with very little variance explained, $F(1, 24) = 0.16$, R^2<.01.

4 Discussion

The present report set out with two primary goals. The first was to conduct a confirmatory analysis demonstrating transfer of AAT to a symbolic arithmetic task, in order to replicate previous work. The second was to conduct an exploratory analysis of specific transfer effects to a battery of untrained ANS-related tasks in order to get a more detailed account of possible training-related changes at the level of the ANS. To varying degrees, we succeeded on both counts.

With respect to the first goal, we observed an approximately 4% increase in accuracy on math from pre- to post-test in the AAT group, while the control group remained stable. The effect size of improvement (standardized gain of AAT group minus control: d=.54) is comparable to previous results using a very similar control condition (d=.47; Park & Brannon, 2013). These effects are arguably small from a practical standpoint (3/80 more problems solved correctly), however, they are impressive in that they are demonstrated with a basic arithmetic task in which our sample of mostly college students have presumably already reached proficiency. Moreover, the improvements are in accuracy and not speed, as reaction time did not significantly differ between groups. This is a novel finding since Park and Brannon's math task was timed, with the dependent variable being the number of items correctly solved in ten minutes, thereby conflating improvements in accuracy with improvements in speed. Our results suggest that the AAT can actually improve calculation proficiency, and not mere processing speed or mental readiness to engage in the task, as has been suggested before (Lindskog & Winman, 2016). In conjunction with Park & Brannon's original reports (2013, 2014) and Park et al. (2016), this is now the fifth experiment to successfully increase math performance with this form of training, suggesting this is a robust effect and further investigation into the underlying mechanisms and properties of AAT is warranted.

Our second goal sought to tackle the question of underlying mechanisms by exploring the extent to which this form of training renders number sense improvements in the processing of both nonsymbolic and symbolic quantities. We found modest preliminary evidence for this in terms of a partial reduction of the distance effect, a classic signature of the ANS (Dehaene et al., 1998). This was demonstrated most promisingly on the Symbolic Comparison task, and to a lesser extent on the Nonsymbolic Comparison task. Nevertheless, we reiterate the exploratory nature of these analyses, and affirm the need for future confirmatory studies to further probe these effects.

With respect to Symbolic Comparison, we observed significant reductions in reaction times in the AAT group under conditions of maximum interference (i.e., trials that were a numerical distance of one away from the reference). This improvement was not observed in any of the other numerical distances, nor was it observed in the control group, suggesting that similar to the math task, improvements were not related to any generalized speeding-up or practice effects at post-test, but were selective to the AAT group and only for the nearest neighbors to the reference. We interpret this selectivity in light of neuroimaging evidence suggesting tuning curve functions in parietal cortex where neurons respond preferentially to a certain quantity, but less so to non-preferred quantities (Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). This implies a certain neuronal population that fires preferentially to the reference number, 23, and a partially overlapping population that fires preferentially to its neighbors, 22 and 24. A putative training-related sharpening of this tuning curve could reduce the neural overlap in representing the signals of neighboring numbers, thus speeding up the processing of these numbers as we observe in our data, but would have relatively little impact on numbers farther from the reference which compete only minimally for representation.

With respect to Nonsymbolic Comparison, we observed that the AAT group outperformed the control group at post-test on two out of the three ratios tested (1.25 and 1.5, but not 1.2). However, we note that the improvements were very small (an advantage of roughly 4% on each ratio), and the post-test difference was driven predominantly by a decrement in the control group, with little improvement from baseline in the AAT group (Figure 5). This drop in performance by the control group might at least partially be explained by the low retest reliability of this task ($r = 0.14$ and 0.43 for the AAT and control groups, respectively; Table S1, SOM). Nevertheless, follow-up regression analyses demonstrate that lower log difference levels achieved during the seven AAT sessions were predictive of greater post-test performance on the Nonsymbolic Comparison task. No such relationship was observed prior to intervention with pre-test performance. Additionally, Fig. 6 shows that the majority of training time was spent training at or around the 1.25 and 1.5 ratios, which showed the largest post-test differences during the Nonsymbolic Comparison task, with relatively little training time spent below those ratios. Therefore, while we hesitate to interpret this as a transfer effect, our combined results do suggest that the training might have some impact on subsequent performance of the Nonsymbolic Comparison task. Altogether, our findings from this study suggest a viable route for AAT to target not only nonsymbolic ANS skills, but also to cross over into the symbolic domain as well to speed up processing during high-interference number comparisons. Such a demonstration, if found to be robust in future studies, is crucial for understanding how approximate arithmetic training on nonsymbolic numerosities may affect the processing and manipulation of the symbolic numbers upon which all math is built.

The present results stand in contrast to previous findings which did not find significant transfer to either a nonsymbolic comparison task or a numeral order judgment task which shares some features with our Symbolic Comparison task (Park & Brannon, 2014). There are several reasons that can explain this discrepancy. First of all, the Park & Brannon study controlled for continuous perceptual cues in their nonsymbolic comparison task, but not during training. Although this is an excellent method for isolating processes related to the extraction of numerical information, it also fundamentally changes the cognitive processes involved between the training and transfer task (c.f., Gebuis & Reynvoet, 2012a; Smets, Moors, & Reynvoet, 2016). Another important distinction to consider is that Park & Brannon tested participants on smaller ratios than used in the present study, with several ratios below 1.2. These tested ratios go well below the typical training range, both in our data as well as their own. Our participants reached average ratios of approximately 1.57 by the *end* of training. Park & Brannon's participants, throughout their three experiments, averaged only slightly better on their last training days, hovering around the 1.5 ratio. Accordingly, despite the overall post-test advantage we observed in the AAT group during our Nonsymbolic Comparison task, we did not see any differences at the lowest ratio, 1.2, at which participants spent relatively little time training (Fig. 6). All these differences notwithstanding, it should be noted that Park & Brannon still found a trend towards improvement on this task.

Although Park & Brannon did not directly measure the symbolic distance effect, they did use a numeral order judgment task that taps a similar process in that it requires organizing a triad of numbers into ascending or descending order as quickly as possible. Although they did find significant improvements in reaction time after AAT, they also found similar effects in one, but not both, of their control training groups, and therefore it is unclear to what extent their effects are simply due to general practice or exposure effects. However, the nature of their task, which involves multiple numerical comparisons with triads of numbers, does not allow them to isolate the effects at a numerical distance of one, which is the only distance in which we saw improvements in our Symbolic Comparison task.

Finally, we admonish that our results must be understood in the context of the psychometric properties of our outcome measures. For instance, there was no systematic correlation among our task battery, indicating that each task may be measuring different aspects of some multi-factorial ensemble of ANS-related skills. This is in agreement with previous literature that shows poor construct validity between different tests assumed to measure the ANS, even different versions of the same test used in different studies (Dietrich et al., 2015; Gilmore et al., 2011; Smets et al., 2014; Smets et al., 2016). Given the disparate nature of these tasks, it is important to plan the selection of transfer tasks carefully in future studies, as it cannot necessarily be assumed that transfer should occur broadly over a number of theoretically related tasks.

Also, we note that there are a couple of tasks that did not convincingly demonstrate all the ANS signatures we were looking for. For example, the Symbolic Estimation task demonstrated unstable coefficients of variation across target sizes, thus violating the assumption of scalar variability, whereas the Symbolic Nonverbal Counting task demonstrated stable error rates across target magnitudes, thus violating the assumption of the size effect. We can only speculate as to the reasons why these tasks failed to demonstrate these hallmark signatures of the ANS. For example, the Symbolic Estimation task generated dots based on a semi-random algorithm, which although meant to dissuade participants from using counting strategies may have also inadvertently added noise to the data that may have distorted coefficient of variation analyses, particularly if participants were not conscientious about refining their answer choices. Potential problems with the Symbolic Nonverbal Counting Task are less clear, but we point out that the range of magnitudes (11–23) was rather narrow, thus possibly minimizing the magnitude of size effects. Whatever the reasons, this speaks further to the inherent difficulty in measuring the ANS, especially when moving away from traditional nonsymbolic comparison tasks. Though difficult, we reiterate the need for such endeavors in future studies looking at transfer of training as comparison tasks alone may not capture the entirety of the transfer effect to the ANS, and do not seem to mediate the effects on symbolic arithmetic.

5 Limitations

One limitation concerns our exploratory analyses, where the primary goal was to generate candidate mechanisms to explain training effects. Due to the non-unitary nature of the ANS, as exemplified by our low inter-task correlations (Table 1), we did not have specific a priori hypotheses as to which, if any, of our tasks would be influenced by training. Therefore, we opted to include a fairly broad set of ANS-related tasks, which opens up issues concerning multiple comparisons. In line with previous recommendations, we decided not to make any corrections for this (Simons et al., 2016). Doing so would filter out all but the strongest effects and hamper the search for these mechanisms, leaving us no closer to understanding the nature of this important effect. Rather, it is more important at this juncture to establish a viable direction for future research to confirm. Our findings suggest the possibility of a partial mitigation of the symbolic, and to a lesser extent nonsymbolic, distance effects, but we caution against over-interpretation at this stage until future confirmatory research is carried out.

Although one of the strengths of our design relative to previous work is that we matched the nonsymbolic stimuli between the training and outcome measures by making all standard visual cues accessible in order to increase process overlap, the fact that we did not control for these continuous variables prohibits any claim of improvement in the extraction of numerical information *per se*. Rather, any interpretation of improvement observed in our Nonsymbolic Comparison task, tenuous as it may be, merely demonstrates that participants were trained to better extract magnitude information from a variety of perceptual cues, including total surface area and density, in addition to numerosity. The extent to which this distinction matters in naturalistic settings, where such perceptual cues are typically confounded anyway (Gebuis & Reynvoet, 2012b), and particularly the extent to which this distinction matters for obtaining improvements in math ability is an important question for future research. This could be addressed, for example, by modifying the training to systematically control for each of these variables in order to decrease reliance on them during training (Fuhs, McNeil, Kelley, O'Rear, & Vilano, 2016), and observe the extent to which such a modified training impacts symbolic arithmetic accuracy. Nevertheless, we note that the strongest transfer effects observed in our study were on symbolic rather than nonsymbolic stimuli. Therefore, we argue that the presence of these continuous perceptual cues throughout our study only minimally impacts our overall results as they are not relevant when processing symbolic stimuli.

Finally, we note that for any given ratio, the approximate arithmetic task is more cognitively challenging than the Nonsymbolic Comparison task because it involves several additional processes (Park & Brannon, 2016) as well as a more elaborate encoding of numerosity. This is problematic because the average ratio that participants reach at the *end* of training (~1.5) is a fairly easy one for most healthy adults in a simple comparison task, leading to ceiling effects in our Nonsymbolic Comparison task on this ratio at pre-test. These ceiling effects may have obstructed our ability to detect change-from-baseline improvements in the AAT group, despite observing a post-test difference between the AAT and control groups. Therefore, a more meaningful demonstration of transfer would require modifying the approximate arithmetic intervention to allow participants to spend more time training at the lower ratios, where more performance variability can be captured by the Nonsymbolic Comparison transfer task.

6 Conclusions

Our results independently replicate Park & Brannon's reports of symbolic arithmetic improvements in young adults after approximate arithmetic training (Park & Brannon, 2013, 2014). We further show that the effects cannot merely be attributed to faster processing at post-test, since both groups improved equally in terms of reaction time, but rather that the AAT group actually improved its calculation accuracy. Moreover, we further sought to better characterize the effects of approximate arithmetic training by demonstrating its influence on the approximate number system. Our preliminary data showed that training performance predicted accuracy on a classic measure of ANS acuity, the Nonsymbolic Comparison task, and despite only very modest improvements after training, the AAT group outperformed the control group at post-test, particularly on discriminating ratios that overlapped more with the training range. Moreover, the fact that we have demonstrated transfer also in the symbolic domain strengthens arguments that the ANS and symbolic number system are integrally linked, and further, that foundational training at the level of the ANS can have downstream effects on higher numerical processing. The distance effect is one of the most salient hallmark features of the ANS (Dehaene et al., 1998), and therefore our demonstration of a partial reduction in this effect after AAT, if found to be robust with future confirmatory research, suggests the precision of the ANS is not static, and may represent a moving target even in adulthood with a relatively short, but targeted intervention. Our work presents a critical step towards understanding the degree to which the ANS is malleable, and offers a potential avenue for future research to explore causal relationships between the ANS and math.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by the MIND Research Institute. J.A and M.B are both employed at the MIND Research Institute, and S.M.J has an indirect financial interest. J.A is also supported by the National Science Foundation Graduate Research Fellowship (Grant No. DGE-1321846), and S.M.J is supported by the National Institute on Aging (Grant No. 1K02AG054665-01).

References

Agrillo C, Piffer L, Bisazza A. Number versus continuous quantity in numerosity judgments by fish. Cognition. 2011; 119(2):281–287. DOI: 10.1016/j.cognition.2010.10.022 [PubMed: 21109238]

Bland JM, Altman DG. Cronbach's alpha. British Medical Journal. 1997; 314(7080):572–572. [PubMed: 9055718]

Buckley PB, Gillman CB. Comparisons of digits and dot patterns. J Exp Psychol. 1974; 103(6):1131–1136. [PubMed: 4457588]

Cappelletti M, Gessaroli E, Hithersay R, Mitolo M, Didino D, Kanai R, Walsh V. Transfer of cognitive training across magnitude dimensions achieved with concurrent brain stimulation of the parietal lobe. J Neurosci. 2013; 33(37):14899–14907. DOI: 10.1523/JNEUROSCI.1692-13.2013 [PubMed: 24027289]

Castronovo J, Gobel SM. Impact of high mathematics education on the number sense. PLoS One. 2012; 7(4):e33832.doi: 10.1371/journal.pone.0033832 [PubMed: 22558077]

Chen QX, Li JG. Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. Acta Psychol (Amst). 2014; 148:163–172. DOI: 10.1016/j.actpsy.2014.01.016 [PubMed: 24583622]

Cordes S, Gelman R, Gallistel CR, Whalen J. Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. Psychon Bull Rev. 2001; 8(4):698–707. doi: DOI: 10.3758/Bf03196206 [PubMed: 11848588]

de Smedt B, Noel M, Gilmore C, Ansari D. How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. Trends in Neuroscience and Education. 2013; 2(2):48–55.

Dehaene S, Dehaene-Lambertz G, Cohen L. Abstract representations of numbers in the animal and human brain. Trends Neurosci. 1998; 21(8):355–361. [PubMed: 9720604]

Dewind NK, Brannon EM. Malleability of the approximate number system: effects of feedback and training. Front Hum Neurosci. 2012; 6:68.doi: 10.3389/fnhum.2012.00068 [PubMed: 22529786]

Dietrich JF, Huber S, Nuerk HC. Methodological aspects to be considered when measuring the approximate number system (ANS) - a research review. Front Psychol. 2015; 6doi: 10.3389/Fpsyg.2015.00295

Dugard P, Todman J. Analysis of Pre-test-Post-test Control Group Designs in Educational Research. Educational Psychology. 1995; 15(2):181–198.

Feigenson L, Libertus ME, Halberda J. Links Between the Intuitive Sense of Number and Formal Mathematics Ability. Child Dev Perspect. 2013; 7(2):74–79. DOI: 10.1111/cdep.12019 [PubMed: 24443651]

Fuhs MW, McNeil NM, Kelley K, O'Rear C, Vilano M. The Role of Non-Numerical Stimulus Features in Approximate Number System Training in Preschoolers from Low-Income Homes. Journal of Cognition and Development. 2016; DOI: doi: 10.1080/15248372.15242015.11105228

Gebuis T, Reynvoet B. The Interplay Between Nonsymbolic Number and Its Continuous Visual Properties. Journal of Experimental Psychology-General. 2012a; 141(4):642–648. DOI: 10.1037/a0026218 [PubMed: 22082115]

Gebuis T, Reynvoet B. The role of visual information in numerosity estimation. PLoS One. 2012b; 7(5):e37426.doi: 10.1371/journal.pone.0037426 [PubMed: 22616007]

Gilmore C, Attridge N, Inglis M. Measuring the approximate number system. Q J Exp Psychol (Hove). 2011; 64(11):2099–2109. DOI: 10.1080/17470218.2011.574710 [PubMed: 21846265]

Halberda J, Feigenson L. Developmental change in the acuity of the "Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. Dev Psychol. 2008; 44(5): 1457–1465. DOI: 10.1037/a0012682 [PubMed: 18793076]

Halberda J, Ly R, Wilmer JB, Naiman DQ, Germine L. Number sense across the lifespan as revealed by a massive Internet-based sample. Proc Natl Acad Sci U S A. 2012; 109(28):11116–11120. DOI: 10.1073/pnas.1200196109 [PubMed: 22733748]

Halberda J, Mazzocco MM, Feigenson L. Individual differences in non-verbal number acuity correlate with maths achievement. Nature. 2008; 455(7213):665–668. DOI: 10.1038/nature07246 [PubMed: 18776888]

Hinrichs JV, Yurko DS, Hu J. Two-Digit Number Comparison: Use of Place Information. Journal of Experimental Psychology-Human Perception and Performance. 1981; 7(4):890–901.

Huck SW, Mclean RA. Using a repeated measures Anova to analyze the data from a pretest-posttest design: A potentially confusing task. Psychol Bull. 1975; 82(4):511–518.

Jaeggi SM, Buschkuehl M, Shah P, Jonides J. The role of individual differences in cognitive training and transfer. Mem Cognit. 2014; 42(3):464–480.

Jaeggi SM, Studer-Luethi B, Buschkuehl M, Su YF, Jonides J, Perrig WJ. The relationship between n-back performance and matrix reasoning - implications for training and transfer. Intelligence. 2010; 38(6):625–635. doi: DOI: 10.1016/j.intell.2010.09.001

Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology. 2013; 49(4):764–766. doi: DOI: 10.1016/j.jesp.2013.03.013

Libertus ME, Feigenson L, Halberda J. Is Approximate Number Precision a Stable Predictor of Math Ability? Learning and Individual Differences. 2013a; 25:126–133. DOI: 10.1016/j.lindif.2013.02.001 [PubMed: 23814453]

Libertus ME, Feigenson L, Halberda J. Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. J Exp Child Psychol. 2013b; 116(4):829–838. DOI: 10.1016/j.jecp.2013.08.003 [PubMed: 24076381]

Libertus ME, Odic D, Halberda J. Intuitive sense of number correlates with math scores on college-entrance examination. Acta Psychol (Amst). 2012; 141(3):373–379. DOI: 10.1016/j.actpsy.2012.09.009 [PubMed: 23098904]

Lindskog M, Winman A. No evidence of learning in non-symbolic numerical tasks - A comment on. Cognition. 2016; 150:243–247. DOI: 10.1016/j.cognition.2016.01.005 [PubMed: 26972468]

Lindskog M, Winman A, Juslin P. Are there rapid feedback effects on Approximate Number System acuity? Front Hum Neurosci. 2013; 7:270.doi: 10.3389/fnhum.2013.00270 [PubMed: 23781191]

Lindskog M, Winman A, Juslin P. The association between higher education and approximate number system acuity. Front Psychol. 2014; 5:462.doi: 10.3389/fpsyg.2014.00462 [PubMed: 24904478]

Lipton JS, Spelke ES. Preschool children's mapping of number words to nonsymbolic numerosities. Child Dev. 2005; 76(5):978–988. DOI: 10.1111/j.1467-8624.2005.00891.x [PubMed: 16149996]

Loosli SV, Buschkuehl M, Perrig WJ, Jaeggi SM. Working memory training improves reading processes in typically developing children. Child Neuropsychology. 2012; 18(1):62–78. DOI: 10.1080/09297049.2011.575772 [PubMed: 21623483]

Lustig C, Shah P, Seidler R, Reuter-Lorenz PA. Aging, training, and the brain: a review and future directions. Neuropsychol Rev. 2009; 19(4):504–522. DOI: 10.1007/s11065-009-9119-9 [PubMed: 19876740]

Mazzocco MM, Feigenson L, Halberda J. Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). Child Dev. 2011; 82(4):1224–1237. DOI: 10.1111/j.1467-8624.2011.01608.x [PubMed: 21679173]

Mohammed S, Flores L, Deveau J, Hoffing RC, Phung C, Parlett CM, Seitz AR. The benefits and challenges of implementing motivational features to boost cognitive training outcome. Journal of Cognitive Enhancement. (in press).

Moyer RS, Landauer TK. Time required for judgements of numerical inequality. Nature. 1967; 215(5109):1519–1520. [PubMed: 6052760]

Mundy E, Gilmore CK. Children's mapping between symbolic and nonsymbolic representations of number. J Exp Child Psychol. 2009; 103(4):490–502. DOI: 10.1016/j.jecp.2009.02.003 [PubMed: 19327782]

Odic D, Hock H, Halberda J. Hysteresis affects approximate number discrimination in young children. J Exp Psychol Gen. 2014; 143(1):255–265. DOI: 10.1037/a0030825 [PubMed: 23163765]

Odic D, Lisboa JV, Eisinger R, Olivera MG, Maiche A, Halberda J. Approximate number and approximate time discrimination each correlate with school math abilities in young children. Acta Psychol (Amst). 2016; 163:17–26. DOI: 10.1016/j.actpsy.2015.10.010 [PubMed: 26587963]

Olsson L, Ostergren R, Traff U. Developmental dyscalculia: A deficit in the approximate number system or an access deficit? Cognitive Development. 2016; 39:154–167. DOI: 10.1016/j.cogdev.2016.04.006

Park J, Bermudez V, Roberts RC, Brannon EM. Non-symbolic approximate arithmetic training improves math performance in preschoolers. J Exp Child Psychol. 2016; 152:278–293. DOI: 10.1016/j.jecp.2016.07.011 [PubMed: 27596808]

Park J, Brannon EM. Training the approximate number system improves math proficiency. Psychol Sci. 2013; 24(10):2013–2019. DOI: 10.1177/0956797613482944 [PubMed: 23921769]

Park J, Brannon EM. Improving arithmetic performance with number sense training: An investigation of underlying mechanism. Cognition. 2014; 133(1):188–200. DOI: 10.1016/j.cognition.2014.06.011 [PubMed: 25044247]

Park J, Brannon EM. How to interpret cognitive training studies: A reply to Lindskog & Winman. Cognition. 2016; 150:247–251. DOI: 10.1016/j.cognition.2016.02.012 [PubMed: 26972469]

Peirce JW. Generating Stimuli for Neuroscience Using PsychoPy. Front Neuroinform. 2009; 2:10.doi: 10.3389/neuro.11.010.2008 [PubMed: 19198666]

Piazza M. Neurocognitive start-up tools for symbolic number representations. Trends in Cognitive Sciences. 2010; 14(12):542–551. DOI: 10.1016/j.tics.2010.09.008 [PubMed: 21055996]

Piazza M, Facoetti A, Trussardi AN, Berteletti I, Conte S, Lucangeli D, Zorzi M. Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. Cognition. 2010; 116(1):33–41. DOI: 10.1016/j.cognition.2010.03.012 [PubMed: 20381023]

Piazza M, Izard V, Pinel P, Le Bihan D, Dehaene S. Tuning curves for approximate numerosity in the human intraparietal sulcus. Neuron. 2004; 44(3):547–555. DOI: 10.1016/j.neuron.2004.10.014 [PubMed: 15504333]

Piazza M, Pica P, Izard V, Spelke ES, Dehaene S. Education enhances the acuity of the nonverbal approximate number system. Psychol Sci. 2013; 24(6):1037–1043. DOI: 10.1177/0956797612464057 [PubMed: 23625879]

Schneider M, Beeres K, Coban L, Merz S, Susan Schmidt S, Stricker J, De Smedt B. Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. Dev Sci. 2016; doi: 10.1111/desc.12372

Schneider M, Siegler RS. Representations of the Magnitudes of Fractions. Journal of Experimental Psychology-Human Perception and Performance. 2010; 36(5):1227–1238. DOI: 10.1037/a0018170 [PubMed: 20873937]

Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, Stine-Morrow EA. Do "Brain-Training" Programs Work? Psychol Sci Public Interest. 2016; 17(3):103–186. DOI: 10.1177/1529100616661983 [PubMed: 27697851]

Smets K, Gebuis T, Defever E, Reynvoet B. Concurrent validity of approximate number sense tasks in adults and children. Acta Psychol (Amst). 2014; 150:120–128. DOI: 10.1016/j.actpsy.2014.05.001 [PubMed: 24875582]

Smets K, Moors P, Reynvoet B. Effects of Presentation Type and Visual Control in Numerosity Discrimination: Implications for Number Processing? Front Psychol. 2016; 7:66.doi: 10.3389/fpsyg.2016.00066 [PubMed: 26869967]

Starr A, Libertus ME, Brannon EM. Number sense in infancy predicts mathematical abilities in childhood. Proc Natl Acad Sci U S A. 2013; 110(45):18116–18120. DOI: 10.1073/pnas.1302751110 [PubMed: 24145427]

StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP; 2013.

Szucs D, Myers T. A critical analysis of design, facts, bias and inference in the approximate number system training literature: A systematic review. Trends in Neuroscience and Education. 2017; 6:187–203.

Szucs D, Nobes A, Devine A, Gabriel FC, Gebuis T. Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. Front Psychol. 2013; 4:444.doi: 10.3389/fpsyg.2013.00444 [PubMed: 23882245]

Wang JJ, Halberda J, Feigenson L. Approximate number sense correlates with math performance in gifted adolescents. Acta Psychol (Amst). 2017; 176:78–84. DOI: 10.1016/j.actpsy.2017.03.014 [PubMed: 28384496]

Wang JJ, Odic D, Halberda J, Feigenson L. Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. J Exp Child Psychol. 2016; 147:82–99. DOI: 10.1016/j.jecp.2016.03.002 [PubMed: 27061668]

Whalen J, Gallistel CR, Gelman R. Nonverbal counting in humans: The psychophysics of number representation. Psychol Sci. 1999; 10(2):130–137. doi: DOI: 10.1111/1467-9280.00120

Wilson AJ, Revkin SK, Cohen D, Cohen L, Dehaene S. An open trial assessment of "The Number Race", an adaptive computer game for remediation of dyscalculia. Behav Brain Funct. 2006; 2:20.doi: 10.1186/1744-9081-2-20 [PubMed: 16734906]

Highlights

- Replicated previously reported transfer of approximate arithmetic training to math
- Math improvements are related to increased accuracy, not speed
- Exploratory results suggest certain number sense skills also improved

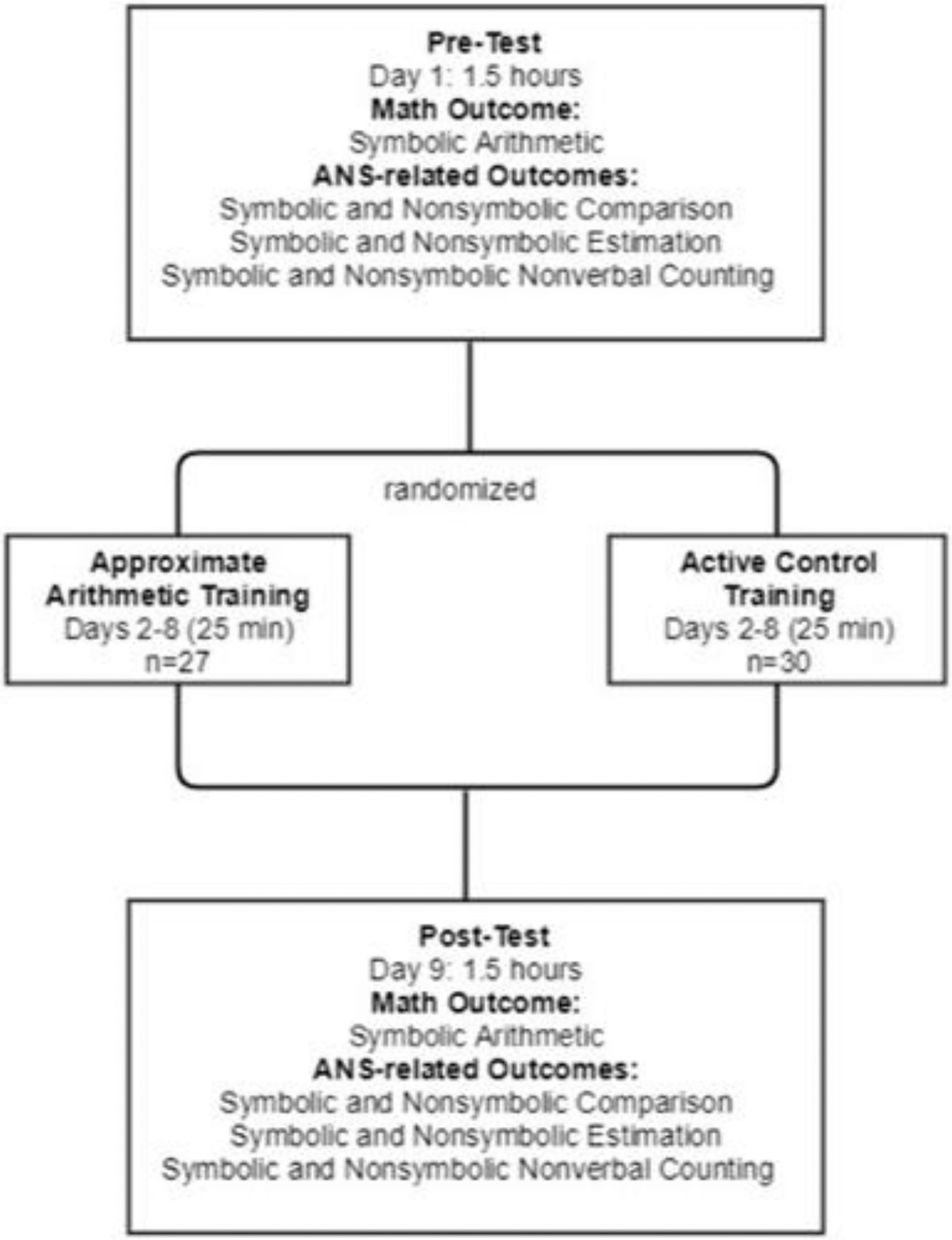
Pre-Test
Day 1: 1.5 hours
Math Outcome:
Symbolic Arithmetic
ANS-related Outcomes:
Symbolic and Nonsymbolic Comparison
Symbolic and Nonsymbolic Estimation
Symbolic and Nonsymbolic Nonverbal Counting
randomized
Approximate Arithmetic Training
Days 2-8 (25 min)
n=27
Active Control Training
Days 2-8 (25 min)
n=30
Post-Test
Day 9: 1.5 hours
Math Outcome:
Symbolic Arithmetic
ANS-related Outcomes:
Symbolic and Nonsymbolic Comparison
Symbolic and Nonsymbolic Estimation
Symbolic and Nonsymbolic Nonverbal Counting

Figure 1. Study Design

The study design consisted of an approximate arithmetic training condition and an active control condition that was bookended by a pre- and post-test.

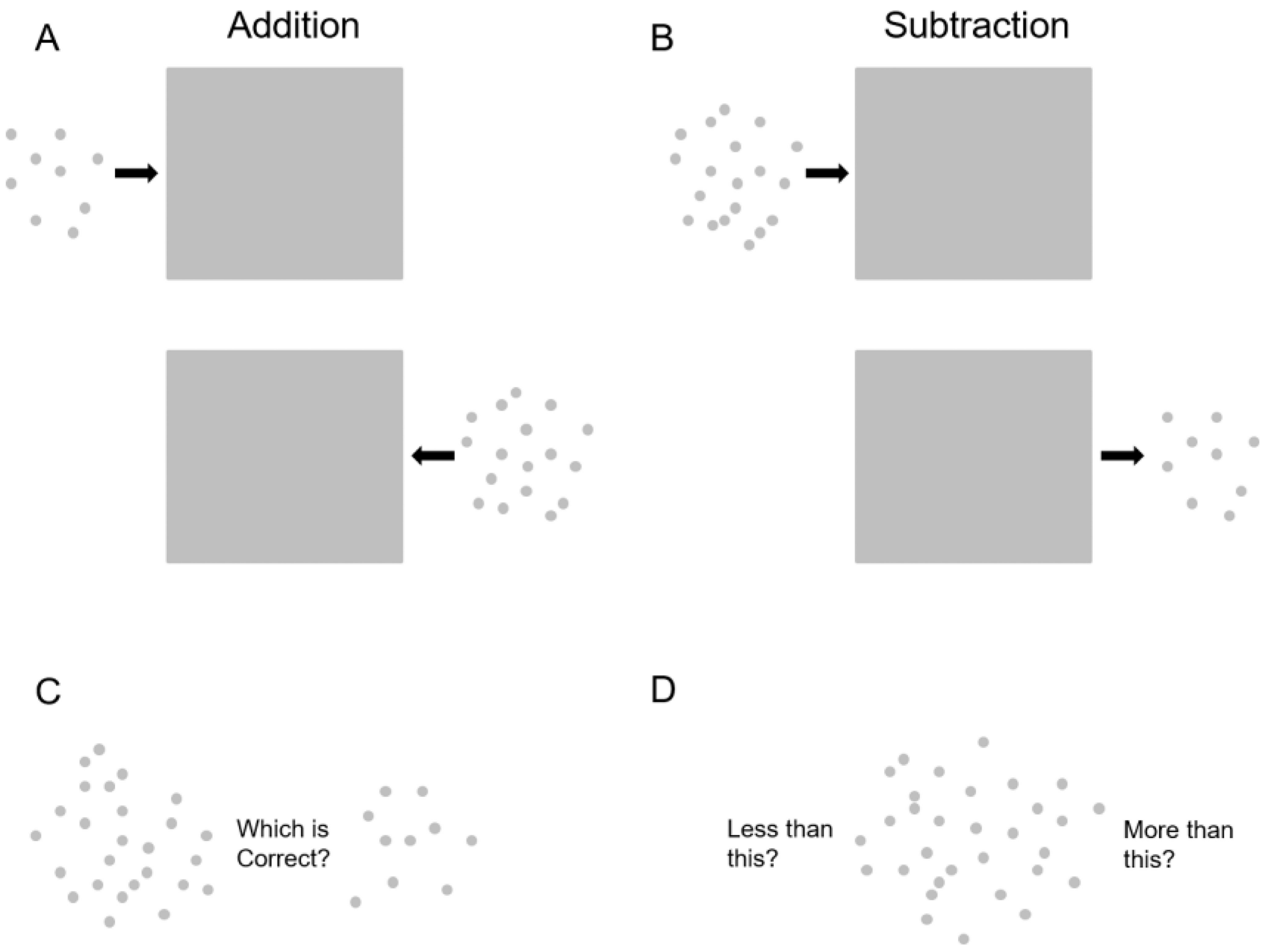

Figure 2. Schematic of Approximate Arithmetic Training

A and B show the two possible trial types, while C and D show the two possible answer formats. A) A dot array appears randomly either to the left or right of the gray square for 1,000ms, before moving behind the square to be occluded. Then another array appears on the contralateral side for 1,000ms before similarly moving into the square. Participant must mentally sum the total number of dots behind the square. B) A dot array appears randomly either to the left or right of the gray square for 1,000ms, before moving into the square to be occluded. Then a smaller array moves out the contralateral side and remains on screen for 1,000ms before disappearing. Participant must mentally track the total number of dots remaining behind the square. C) One possible answer format is shown, where the participant must choose the correct answer from one of two possible choices using either the “A” or “L” keys to indicate the left or right choice, respectively. D) Another possible answer format is shown, where the participant must decide whether the correct answer is less or more numerous than the reference display, by pressing “A” or “L”, respectively.

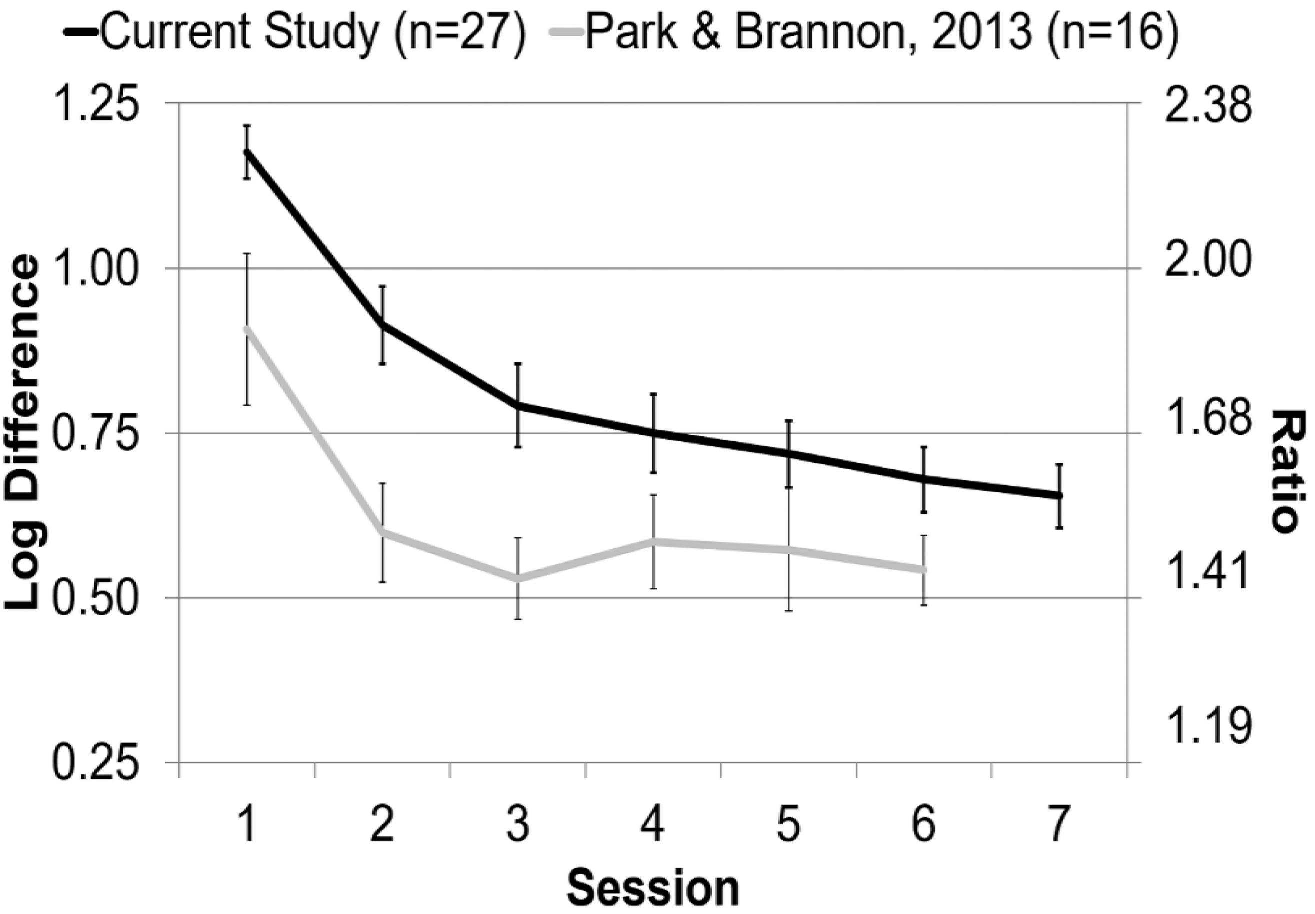

Figure 3. Approximate Arithmetic Training Curves

Training results show significant improvement over time, both in our data as well as Park & Brannon’s original results using a similar training task. Y-axis on the left represents the log difference level, while the axis on the right represents their ratio conversion. For example, on Session 1, our training group on average were comparing the correct answer to an array of dots that was approximately 2.25 times greater or smaller. Error bars represent standard error.

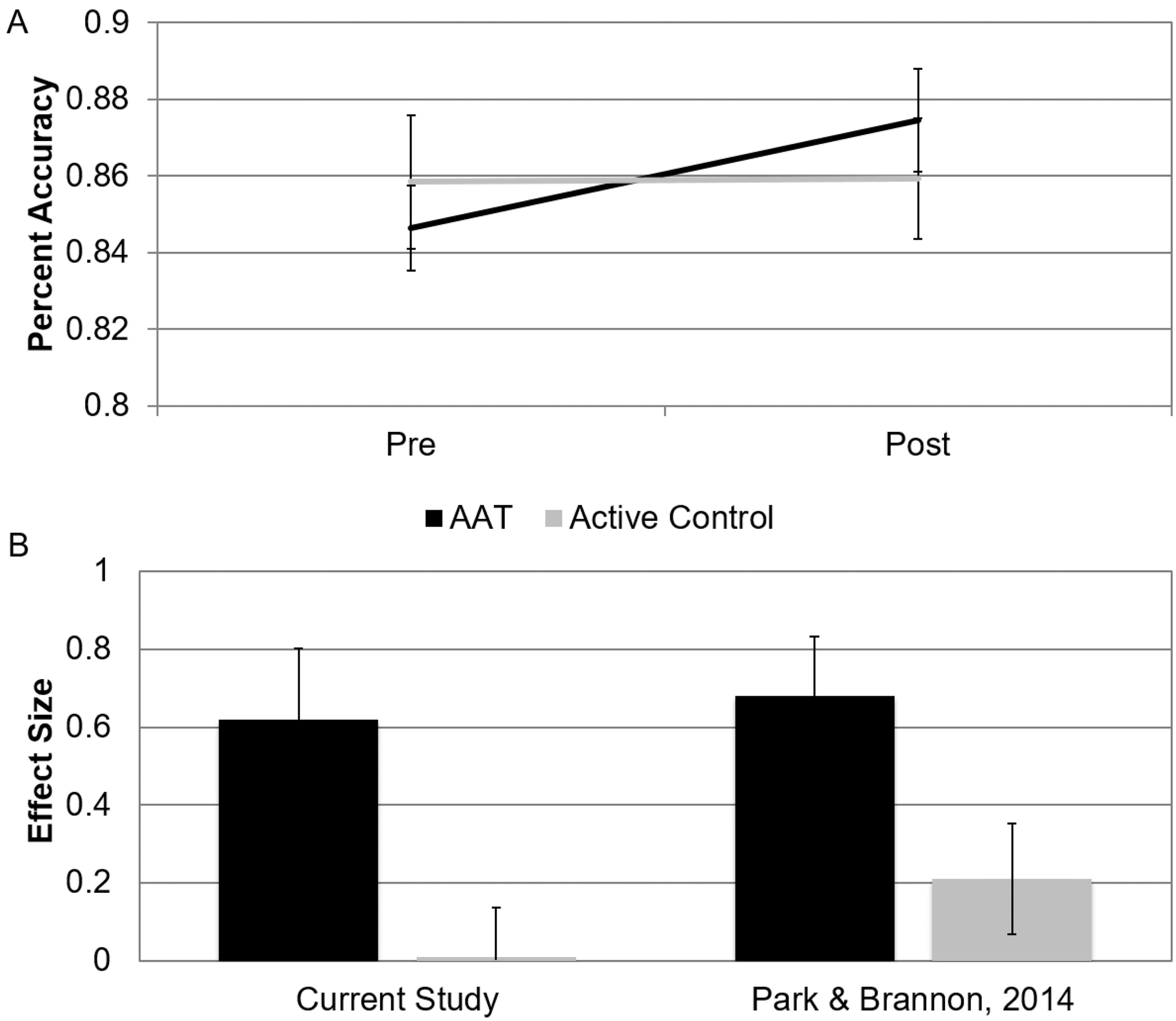

Figure 4. Symbolic Arithmetic Task

Improvements in symbolic arithmetic are observed after training for the AAT group only, both in the current study as well as in Park & Brannon's original report. Error bars represent standard error.

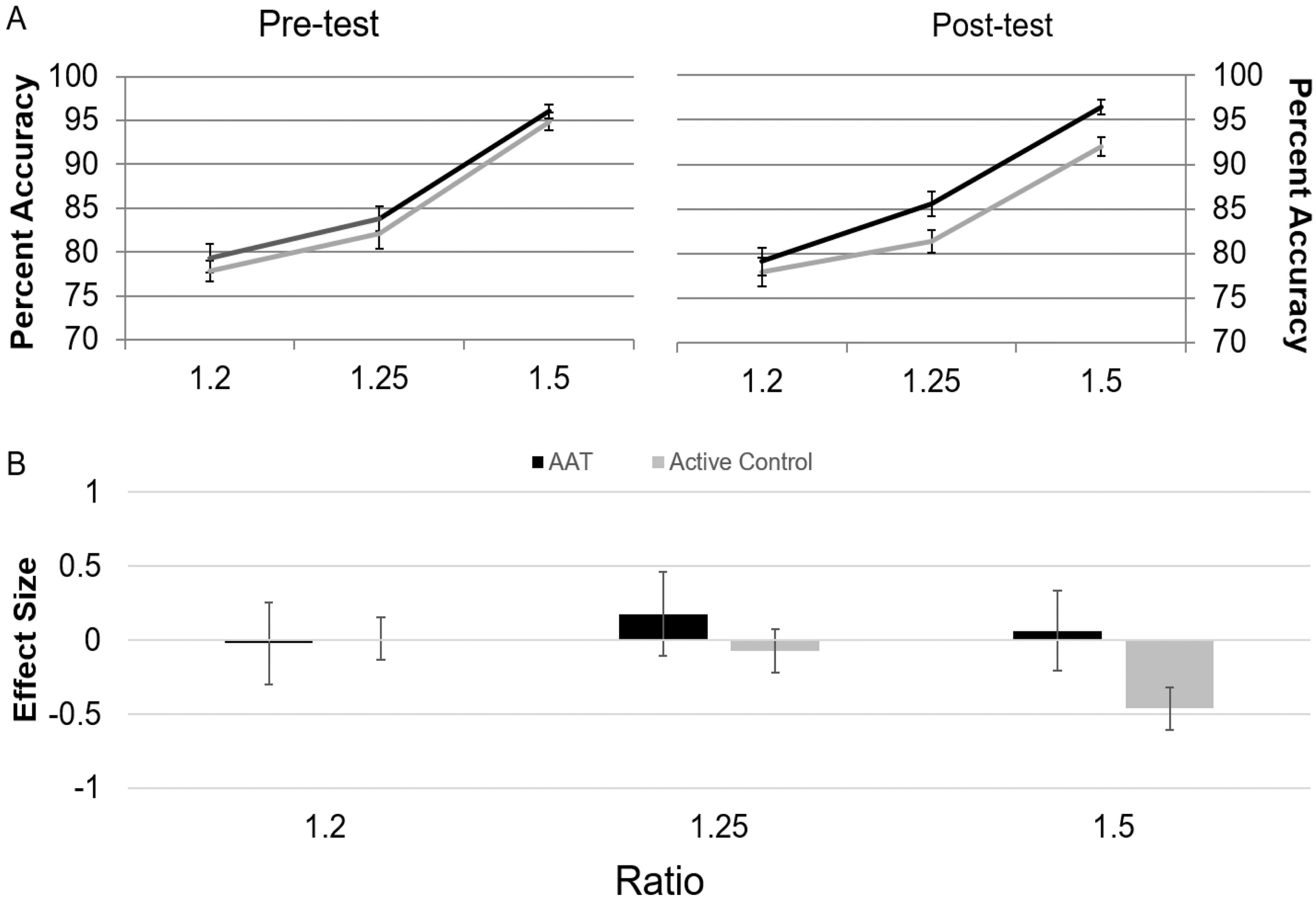

Figure 5. Nonsymbolic Comparison Task

The AAT group outperformed the control group at post-test, with differences driven primarily by the 1.25 and 1.5 ratios. However, this effect is driven mostly by worsening performance in the control group rather than any improvements in the AAT group. Error bars represent standard error.

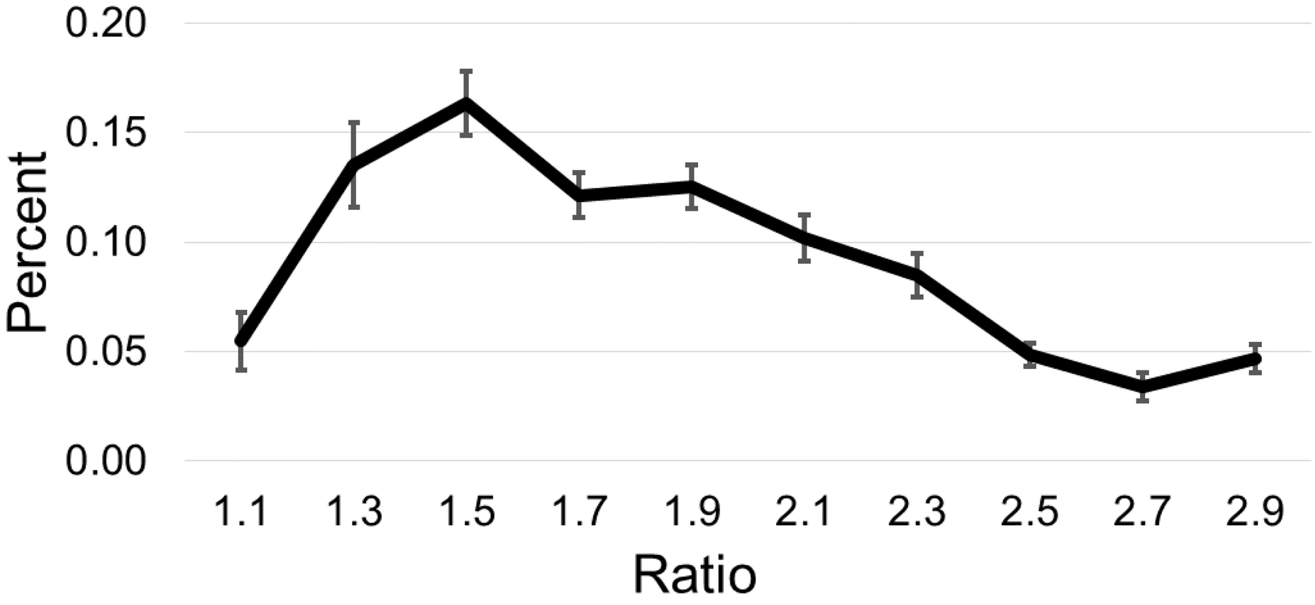

Figure 6. Time on Task

Proportion of time spent training on different ratios is graphed over all 7 sessions of the AAT. Ratios were binned into deciles, with each decile spanning a range of 0.2. The x-axis represents the average ratio of each binned decile. For example, the first bin spans from 1 to 1.2, and the next one from 1.2 to 1.4, etc. The majority of training time was spent at and around a ratio of 1.5, with only about 6% of time spent training at a ratio of 1.2 or less.

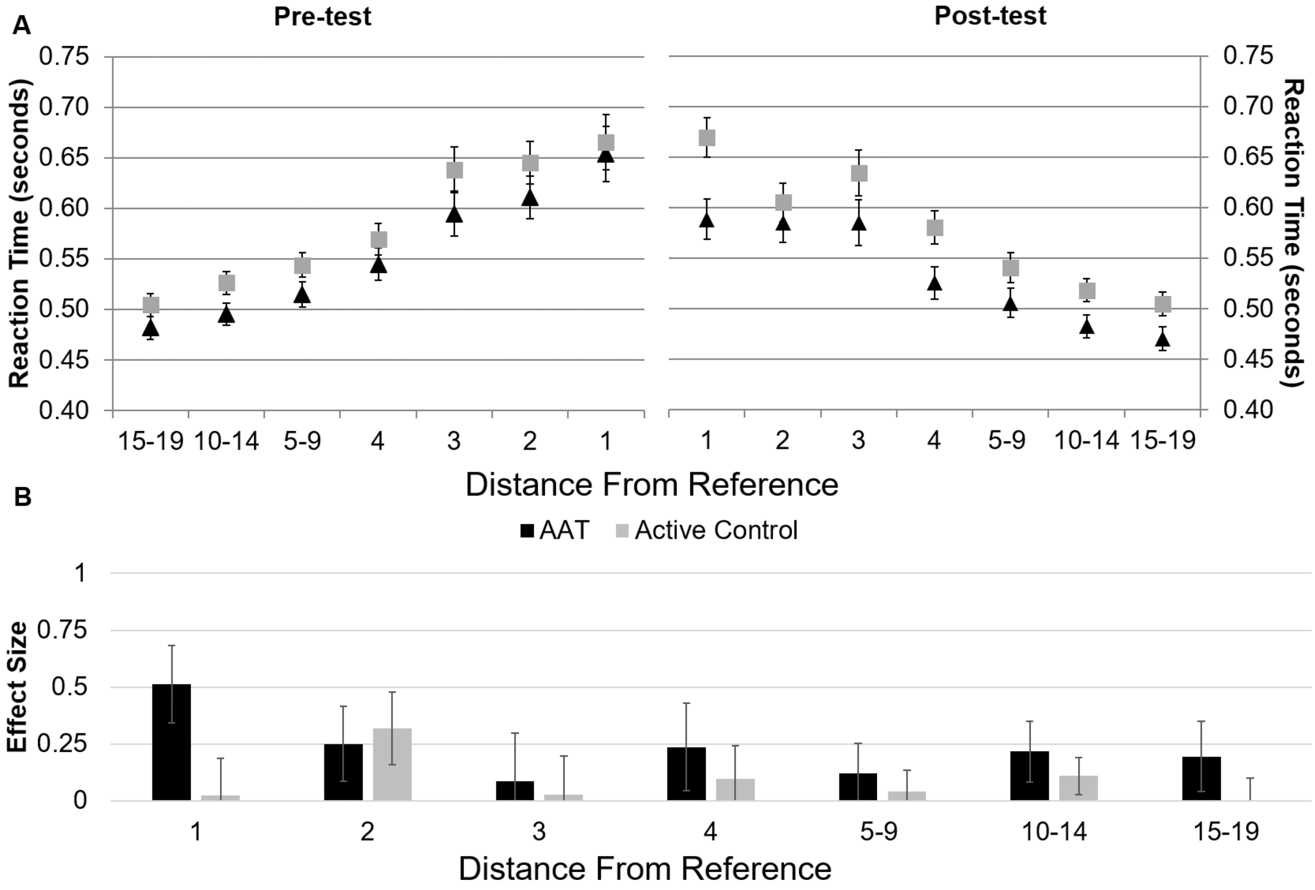

Figure 7. Symbolic Comparison

A) The signature distance effect of increasing reaction time with decreasing numerical distance to the reference (23) is apparent in both the ANS and control groups. More importantly, at post-test, the AAT group showed significantly faster processing during trials with the highest interference (at the ±1 distance), relative to the control group. B) Effect sizes are shown for each distance bin for both groups.

Table 1

Correlation matrix of all tasks performed at pre-test and the first approximate arithmetic training session.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1 – Nonsymbolic Comparison	–							
2 – Symbolic Comparison	−.04	–						
3 – Fraction Comparison	**.31**	.15	–					
4 – Nonsymbolic Estimation	.08	.00	.21	–				
5 – Symbolic Estimation	.23	.13	−.03	−.01	–			
6 – Nonsymbolic Nonverbal Count	−.08	.14	.06	−.08	.01	–		
7 – Symbolic Nonverbal Count	.18	.03	.24	.11	.24	**.57**	–	
8 – Symbolic Arithmetic	.14	−.05	−.01	.07	−.13	.01	−.01	–
9 – Training Session 1	.06	.19	**.43**	**.50**	.06	.27	.09	.15

Note. Bolded items are significant at p < .05. All signs are flipped where appropriate so that positive correlations reflect positive relationships (i.e., mutually good performance) between tasks. In general, N=57 for all correlations, except those involving Training Session 1, which has n=27.

Table 2

Regression of Training Performance on Transfer.

	Symbolic Arithmetic				**Symbolic Comparison**				**Nonsymbolic Comparison**			
	b	**SE**	**β**	**R^2**	**b**	**SE**	**β**	**R^2**	**b**	**SE**	**β**	**R^2**
Intercept	.25	.18			.12	.07			**.81**	**.14**		
Training Performance	.02	.05	.07	.01	.01	.04	.04	.03	**.11**	**.03**	**.55**	**.29**
Pre-test Covariate	**.73**	**.20**	**.62**		**.73**	**.12**	**.77**		.17	.16	.19	

Note: b represents the unstandardized coefficient, while *β* represents the standardized beta coefficient. Significant values ($p<.05$) are displayed in boldface. R^2 values are calculated based only on Training Performance (i.e., pre-test covariate is removed from the model) in order to prevent undue inflation due to the contributions of pre-test performance to the total explanatory power of the model. All signs are flipped when appropriate so that positive coefficients denote a relationship between greater training performance and greater transfer performance.