JOURNAL OF COGNITION AND DEVELOPMENT, 17(5):737–764

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2015.1105228

The Role of Non-Numerical Stimulus Features in Approximate Number System Training in Preschoolers from Low-Income Homes

Mary Wagner Fuhs
University of Dayton

Nicole M. McNeil, Ken Kelley, Connor O'Rear, and Michael Villano
University of Notre Dame

Recent findings have suggested that adults' and children's approximate number system (ANS) acuity may be malleable through training, but research on ANS acuity has largely been conducted with adults and children who are from middle- to high-income homes. We conducted 2 experiments to test the malleability of ANS acuity in preschool-aged children from low-income homes and to test how non-numerical stimulus features affected performance. In Experiment 1, mixed-effects models indicated that children significantly improved their ratio achieved across training. Children's change in probability of responding correctly across sessions was qualified by an interaction with surface area features of the arrays such that children improved their probability of answering correctly across sessions on trials in which numerosity conflicted with the total surface area of object sets significantly more than on trials in which total surface area positively correlated with numerosity. In Experiment 2, we found that children who completed ANS acuity training performed better on an ANS acuity task compared with children in a control group, but they only did so on ANS acuity trials in which numerosity conflicted with the total surface area of object sets. These findings suggest that training affects ANS acuity in children from low-income homes by fostering an ability to focus on numerosity in the face of conflicting non-numerical stimulus features.

Understanding the foundation of children's mathematical knowledge is crucial for developing high-quality instruction and early interventions, particularly for young children from low-income homes, who are at risk for poor mathematics achievement. One potentially foundational system that has been the focus of recent research is the approximate number system (ANS). This system is thought to underlie young children's ability to estimate and compare large quantities without counting. Recent evidence has suggested that the acuity of the ANS may be related to mathematics achievement (e.g., Halberda, Mazzocco, & Feigenson, 2008; Libertus, Feigenson, & Halberda, 2011), is malleable (e.g., DeWind & Brannon, 2012), and can be improved by childhood interventions (e.g., Hyde, Khanum, & Spelke, 2014). However, previous research on the development and training of ANS

Address correspondence to Mary Wagner Fuhs, Department of Psychology, University of Dayton, SJ 319, 300 College Park, Dayton, OH 45469, USA. E-mail: mfuhs1@udayton.edu
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hjcd.

acuity has focused almost exclusively on participants from middle- to high-income homes, which leaves open questions about the generalizability to young children living in low-income homes, who may stand to benefit most from such training. In two experiments, we tested if an adaptive ANS acuity training program influences the ANS acuity of young children from low-income homes and whether or not ANS acuity is affected by non-numerical stimulus features.

Variations in ANS Acuity

The ANS provides individuals with the ability to estimate and compare numerical quantities without counting (Feigenson, Dehaene, & Spelke, 2004). The acuity of this system depends at least in part on Weber's law of "just noticeable difference," or the idea that at a certain ratio threshold, one can reliably perceive a difference between the numerosities of two nonsymbolic quantities (e.g., two sets of dots, shapes). Detecting smaller differences in ratios improves across development, with particularly rapid improvements in early childhood. When presented with two arrays simultaneously, preschoolers have been shown to be capable of discriminating arrays at a 3:2 or 4:3 ratio (Halberda & Feigenson, 2008), and adults have been shown to be capable of discriminating arrays at an 8:7 ratio (or even more precisely for some individuals; Halberda & Feigenson, 2008; also see Barth, Kanwisher, & Spelke, 2003). ANS acuity not only improves dramatically across development, but also varies widely between individuals within the same age group (e.g., Halberda & Feigenson, 2008). For example, Halberda, Ly, Wilmer, Naiman, and Germine (2012) reported both developmental variability and individual variability within same-age groups in ANS acuity in a large-scale Internet sample of participants ranging in age from 11 years to 85 years old. According to their findings, there is such wide individual variation within age groups that "one adult in eight has a number sense that is less precise than a typical 11-y-old child's" (Halberda et al., 2012, p. 4).

There are competing explanations for these intraindividual and interindividual differences in ANS acuity. Some researchers argue that ANS acuity is separate from perceptual properties of stimuli such that variations in the preciseness of an individual's ANS and the ratio between numerosities of object sets alone can account for the differences (Dehaene, 2011; Stoianov & Zorzi, 2012). However, there is increasing evidence to support the idea that continuous stimulus features (e.g., surface area, convex hull) influence and may even be inseparable from number representations (Clayton & Gilmore, 2015; Gebuis & Reynvoet, 2011; Lourenco & Longo, 2010; Walsh, 2003). More specifically, the competing-processes account (Clayton & Gilmore, 2015) suggests that intraindividual and interindividual differences in ANS acuity are in part the result of the ability to take advantage of helpful stimulus features that are correlated with numerosity and inhibit irrelevant but salient features, such that performance on ANS acuity tasks is the result of numerical processing, visual processing, and inhibitory control processing. These differing accounts have implications for how we interpret research findings on the associations between ANS acuity and mathematics achievement.

ANS Acuity and Mathematics Achievement

Some researchers have argued that children initially learn symbolic mathematical concepts by mapping them onto their preexisting ANS (e.g., Dehaene, 2011; Feigenson et al., 2004; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005). Findings from several studies support this argument as they have shown a unique association between ANS acuity and symbolic mathematics ability in school-aged children (e.g., De Smedt, Verschaffel, & Ghesquière, 2009; Halberda et al., 2008; Inglis, Attridge, Batchelor, & Gilmore, 2011; Keller & Libertus, 2015). This association may also hold for younger children who have little exposure to formal schooling, which opens up the possibility that children's ANS acuity is foundational for learning symbolic mathematics (e.g., Libertus et al., 2011; Mazzocco, Halberda, & Feigenson, 2011; Starr, Libertus, & Brannon, 2013). However, the studies conducted in this area have generally been correlational and have produced mixed findings, including reports that ANS and mathematics achievement are uncorrelated or that correlations depend on shared variance with inhibitory control, thereby suggesting support for the competing-processes account of ANS acuity (e.g., Bonny & Lourenco, 2013; Fuhs & McNeil, 2013; Gilmore et al., 2013; Holloway & Ansari, 2009; Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013; Sasanguie, Defever, Maertens, & Reynvoet, 2014). Given the problems with using correlational analyses of longitudinal data sets to make conclusions about the potential benefits of training specific early mathematics skills (Bailey, Watts, Littlefield, & Geary, 2014), it is unclear if ANS acuity training is a viable option for early training for children who are at risk for mathematics difficulties. In fact, no study to date has even shown that ANS acuity training can improve the ANS acuity of young children from low-income homes, let alone their mathematics achievement.

Several lines of research do suggest a role for experiential variables in the precision of children's nonsymbolic ANS acuity. For example, Nys et al. (2013) found evidence that mathematics education sharpened nonsymbolic ANS acuity. Park and Brannon (2013, 2014) found that nonsymbolic approximate arithmetic training improved ANS acuity in adults. Hyde and colleagues (2014) found that a short ANS acuity intervention improved children's performance on exact numerosity tasks. Unfortunately, much of the knowledge about ANS acuity to date has come from studies of children from middle- to high-income homes. The few studies that have been conducted with young children from low-income homes have shown much better performance on assessments of approximate addition for children from middle- to high-income homes compared with children from low-income homes (McNeil, Fuhs, Keultjes, & Gibson, 2011) and weaker associations between ANS acuity and mathematics achievement than those found in previous studies (Fuhs & McNeil, 2013). Research has consistently shown that preschoolers from low-income homes are more likely to struggle with early mathematics concepts and to have less exposure to high-quality mathematics materials in their environment (e.g., Ramani & Siegler, 2008). At the same time, there is promising evidence to suggest that instructional programs can boost the general mathematics skills of children from working-class and low-income homes (e.g., Clements & Sarama, 2007; Dyson, Jordan, & Glutting, 2013; Klein, Starkey, Clements, Sarama, & Iyer, 2008; Ramani & Siegler, 2008; Whyte & Bull, 2008; Wilson, Dehaene, Dubois, & Fayol, 2009). Thus, it is important to address the generalizability of ANS acuity training programs to children from low-income homes because they stand to benefit most from such programs.

How Do Improvements in ANS Acuity Occur?

Beyond addressing if ANS acuity training programs can improve the ANS acuity of children from low-income homes, it is also important to understand specific mechanisms that may explain potential improvements. In particular, do improvements in ANS acuity depend on the ability to inhibit attention to salient non-numerical stimulus features, such as the specific types of individual objects used in the training, or the extent to which irrelevant stimulus features (e.g., total surface area) conflict with the numerosity of the sets being compared? Answers to such questions are essential for both theory and practice because they inform our understanding of the specific processes through which improvements in ANS acuity can occur.

In terms of the specific types of individual objects used in the training, some objects (e.g., animals or toys) may be more attention-grabbing and/or distracting than others (e.g., solid-colored black dots). Researchers have debated the use of perceptually rich, concrete stimuli versus idealized stimuli to teach mathematics and science concepts (Goldstone & Sakamoto, 2003; Kaminski, Sloutsky, & Heckler, 2008; McNeil, Uttal, Jarvin, & Sternberg, 2009). Some have argued that concrete stimuli are better because they capture children's attention and help them stay focused on the task at hand (Burns, 1996). Others have argued that this attention-grabbing aspect of concrete objects is precisely why it is more beneficial to use idealized stimuli when teaching math concepts to children, as idealized objects are stripped of their perceptual details, are less likely to distract children from the mathematics at hand (compared with using concrete stimuli), and are more likely to help children construct knowledge that is transferrable to other, related tasks (Kaminski et al., 2008; Sloutsky, Kaminski, & Heckler, 2005). There may be a trade-off, however, between benefits of concrete and idealized representations depending on other factors, such as the complexity of the mathematics task (Koedinger, Alibali, & Nathan, 2008) and the amount of irrelevant prior knowledge children have associated with the objects (Petersen & McNeil, 2013). It is currently unclear if the specific objects used in the to-be-compared arrays influence children's performance on ANS acuity tasks, particularly during training. In Halberda and Feigenson's (2008) original ANS paradigm, children were asked to compare sets of familiar concrete objects. However, other researchers have presented ANS tasks using more idealized stimuli, such as tones or solid-colored dots (e.g., Barth, La Mont, Lipton, & Spelke, 2005).

Another way in which non-numerical stimulus features may play a role in children's ANS acuity is the extent to which irrelevant stimulus features conflict with the numerosity of the sets. Typical ANS acuity tasks often include several different trial types to control for features like surface area. Recent evidence has suggested that surface area features of ANS acuity stimuli may influence young children's performance, such that they are more likely to get ANS task trials correct when surface area is positively correlated with numerosity (Fuhs & McNeil, 2013; Gilmore et al., 2013; Soltész, Szűcs, & Szűcs, 2010). Thus, children's ANS acuity may be inflated by trials in which surface area is correlated with numerosity. If children are less sensitive to discrete stimuli because of their young age or because of not being sufficiently exposed to numerical information in early development, we might expect them to need to rely on inhibitory control processes for ANS acuity trials in which numerosity of object sets conflicts with other stimulus features (e.g., total surface area of object sets). Thus, ANS acuity training may help children learn to focus on numerosity in the face of conflicting features of stimuli, a skill that would certainly improve performance on ANS acuity tasks, particularly on trials in which irrelevant stimulus features are particularly salient. Evidence exists to suggest that improvements in ANS acuity may depend on other non-numerical cognitive processes needed to disentangle numerosity from other stimulus features. For example, DeWind and Brannon (2012) showed that improvements in adults' ANS acuity were driven by a decrease in the bias to respond based on surface area cues.

We conducted two experiments to examine if ANS acuity training influences ANS acuity in preschoolers from low-income homes. In both experiments, we designed the stimuli to test the extent to which non-numerical features influence children's responses to training. In the first experiment, we exposed preschoolers to a computerized adaptive training paradigm to test if children progressed to more difficult ratios across training sessions (evidence of training effects on ANS processing) and if children's accuracy depended on the non-numerical features of the individual objects or sets. We hypothesized that children would progress to more difficult ratios across training and that changes in accuracy across training would depend on varying stimulus features of the individual objects and sets. We did not have a directional hypothesis for the effect of the individual objects used in the training because of the conflicting predictions in the literature. However, we did have a directional hypothesis related to the trials in which surface area conflicts with the numerosity of the sets. Specifically, we predicted that children would improve more on trials in which numerosity conflicted with surface area features of object sets compared with trials in which numerosity was positively correlated with the total surface area of object sets. In the second experiment, we extended the results of the first experiment by conducting a randomized experiment that included a control group and a pretest and posttest measure of ANS acuity that was separate from the training trials themselves. Here we hypothesized that children who completed the adaptive ANS acuity training would exhibit better ANS acuity at posttest compared with children in the control group, particularly for the trials in which surface area of the object sets conflicted with numerosity.

EXPERIMENT 1

Method

Participants. Approval for this project was obtained from a university Human Subjects Review Board as well as from the Head Start program from which the participants were drawn. Head Start promotes school readiness of children under 5 from low-income families through education, health, social and other services" (Office of Head Start, n.d., par. 1). Preschoolers whose families meet the U.S. federal poverty guidelines are eligible for participation in Head Start. Thirty children (50% girls) completed the adaptive computerized ANS acuity training. There were eight boys and eight girls in the black-dots stimulus condition and seven boys and seven girls in the colorful objects stimulus condition. Children were aged, on average, 4 years 9 months at the beginning of training (SD = 5 months). Although the precise race and ethnicity makeup of the sample was not available, the demographics of the programs were 33% White, 33% African American or Black, and 33% Hispanic.

Experimental ANS Training Task. We designed an adaptive computer program to train children in ANS acuity. Children were presented with two arrays of objects on the screen, each enclosed in a rectangular box, and were asked to choose which side had a greater number of objects. The computer program was developed in Java using video game programming techniques (e.g., Davison, 2005). The program ran in full-screen, exclusive mode at a resolution of 1,280 pixels × 800 pixels. Two black, framing rectangles against a white background were drawn on the screen at all times and served as a fixation point when stimuli were not present. Each rectangle measured 508 pixels × 604 pixels. Individual stimuli, either dots or objects, were created in five different image sizes to accommodate the different control types for surface area (described below). Each image was created on a white background that was not discernible against the white background of the rest of the display. To display the images within the rectangles in a random pattern, each rectangle was internally represented within the program as either a grid of 16 elements (four rows by four columns) or a grid of 30 elements (six rows and five columns). In the 16-element grid, each cell was defined as 120 pixels wide by 125 pixels high, and this grid was utilized if any of the images to be displayed were of the largest size. Otherwise, if none of the largest images were to be presented, the 30-element grid was used and was composed of 100-pixel × 100-pixel cells. To display an individual image on the screen, a cell location was randomly selected and then, the image was randomly located entirely within the borders of one of these cells of the grid. This grid technique ensured there would be no overlapping of individual images, but the random placement within a cell reduced the overall uniformity of the entire display.

Children were randomly assigned to either the dots or objects condition, with half of the sample receiving training with all objects as black dots and the other half receiving training with colorful pictures of animals (e.g., pigs, cows) and child-friendly inanimate objects (e.g., shoes, balls; see Figure 1 for example computer screenshots).

Children wore noise-cancelling headphones during training to receive verbal prompts and feedback. A local first-grade teacher with theatrical experience recorded the verbal prompts and feedback for the computer program, so the program would be as child friendly as possible. On each trial, children were verbally prompted by the teacher to turn their attention to the computer screen (e.g., “OK. Ready for the next one?”). Following the prompt, the stimulus appeared on the screen for 1,500 ms along with another verbal prompt, “Which one has *more* dots (or pigs, shoes, etc. for the colorful objects condition)?” A stimulus duration of 1,500 ms was chosen as this duration allowed children time to view the stimulus but did not allow enough time for counting. Children were given an unlimited amount of time to respond after the stimulus disappeared from the screen. Children indicated their responses by pressing a computer key that corresponded to the side of the box with more dots or colorful objects (keys had red stickers on them to indicate the appropriate keys on the keyboard). Following children’s responses, feedback was provided. If the child was correct, he or she heard a ding that signaled a correct response along with verbal feedback consisting of one of three randomly varying phrases, “That’s right!” “That’s correct; great job!” and “You got it; way to go!” If the child was incorrect, he or she heard a buzzer sound along with verbal feedback consisting of one of three randomly varying phrases: “That’s not right, but you can try again”; “Whoops! You missed that one; try again”; and “Good try, but that’s not it; please try again.”

This computer training was adaptive such that it proceeded to the next most difficult ratio after children got three consecutive correct answers. All children began training at a 4:1 ratio, and the training was adaptive across sessions, such that if children ended Session 1 at a 4:3 ratio, they began Session 2 at a 4:3 ratio. If a child answered incorrectly, then the program went back to the closest easier ratio. For example, if a child answered a 4:3 ratio trial incorrectly, then his or her next trial was a 3:2 ratio trial. There were 10 total levels of ratio possibilities built into the program, 4:1, 2:1, 3:2, 4:3, 5:4, 6:5, 7:6, 8:7, 9:8, and 10:9. For each

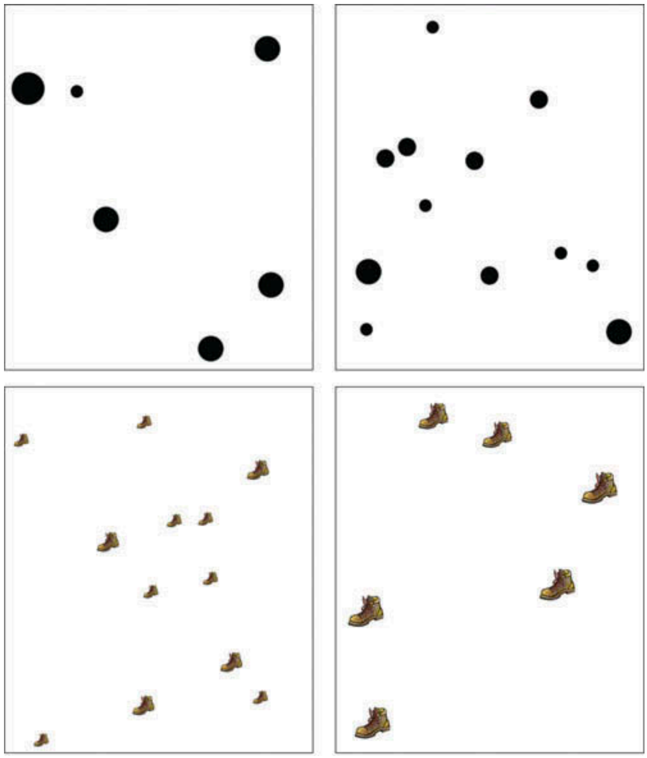

FIGURE 1 Example screenshots of trials in the dots condition and the colorful objects condition comparing items at a 1:2 ratio with surface area inversely related to numerosity.

ratio, the specific numbers of dots (or objects) in the two boxes were randomly selected from among 15 possibilities, 5 for each control type. The number of dots in each box ranged from 1 to 30 dots.

We used three different control types to test whether or not, or to what extent, children's responses to training were influenced by numerically irrelevant stimulus features. Specifically, three types of trials were randomly presented to children: a) *mean area equal* trials in which the mean surface area (mean dot size) was equated such that the larger object set also had more total surface area; b) *total area equal* trials in which the total surface area was equated across object sets; and c) *inverse* trials such that the numerically smaller object set had more surface area at an inverse ratio (e.g., if numerosity varied between the displays at a 4:1 ratio, surface area varied at a 1:4 ratio). These control types have been used previously in ANS acuity tasks for young children (e.g., Fuhs & McNeil, 2013; Gilmore et al., 2013; Halberda & Feigenson, 2008; Libertus et al., 2011). After every 10 trials, children heard a verbal prompt that it was time for a sticker break. Children were then allowed to pick a sticker to put on their sticker page. At the end of the eighth and final training session, children were able to take their sticker pages filled with stickers back to their classrooms.

Training Procedure. All children participated in eight 10-min ANS training sessions (two sessions per week) while an experimenter observed the child to ensure the testing protocol was followed and provided encouragement and assistance to the child as needed. A total training time of 80 min was chosen as improvement was seen in mathematics skills in previous studies after 80 min of instruction (Ramani & Siegler, 2008; Whyte & Bull, 2008). All children completed training in a quiet area of their school.

Results

Descriptive Statistics. Table 1 displays the number of trials each child completed, the number of correct trials, and the mean ratio of the trials within each session. Across sessions, children achieved an average ratio ranging from 4:1 to 2:1.

Analytic Plan. We used mixed-effects (multilevel) models to test our research questions of interest. Mixed-effects analyses were conducted in the R language and environment for statistical computing (R Development Core Team, 2012) using the lme4 (Bates, Maechler, & Bolker, 2011) package. We used mixed-effects modeling due to the flexibility of the model and its ability

TABLE 1
Experiment 1 descriptive statistics within sessions

Session		N	*Minimum*	*Maximum*	*M*	*SD*
1	Total Trials	30	17.00	43.00	31.33	5.77
	Total Correct Trials	30	10.00	32.00	22.57	7.48
	Average Ratio Achieved	30	1.89	4.00	2.96	0.75
2	Total Trials	30	15.00	44.00	34.10	6.24
	Total Correct Trials	30	6.00	35.00	23.47	7.61
	Average Ratio Achieved	30	1.30	4.00	2.66	0.90
3	Total Trials	30	10.00	43.00	33.80	7.49
	Total Correct Trials	30	7.00	33.00	23.60	7.14
	Average Ratio Achieved	30	1.20	3.95	2.64	0.94
4	Total Trials	30	17.00	44.00	33.43	5.64
	Total Correct Trials	30	10.00	33.00	23.87	6.22
	Average Ratio Achieved	30	1.33	4.00	2.67	0.83
5	Total Trials	30	17.00	42.00	32.93	4.68
	Total Correct Trials	30	7.00	32.00	22.73	5.50
	Average Ratio Achieved	30	1.26	4.00	2.69	0.85
6	Total Trials	30	16.00	36.00	30.73	4.58
	Total Correct Trials	30	11.00	30.00	21.20	5.81
	Average Ratio Achieved	30	1.21	4.00	2.74	0.90
7	Total Trials	30	13.00	39.00	30.63	6.20
	Total Correct Trials	30	5.00	31.00	21.90	6.39
	Average Ratio Achieved	30	1.39	4.00	2.58	0.83
8	Total Trials	30	13.00	40.00	30.80	5.65
	Total Correct Trials	30	6.00	32.00	20.80	7.13
	Average Ratio Achieved	30	1.23	4.00	2.80	0.88

Note. For average ratio achieved, smaller numbers indicate *better* performance.

to address our specific research questions and to appropriately address the nested structure of the repeated-measures data. We first examined if children advanced to more difficult ratios across training. We used a linear mixed-effects model for the continuous outcome of ratio (as a decimal) on each trial (smaller ratio indicates better performance) and used a model comparison approach to select between competing models of change across trials and sessions. To address whether or not and the degree to which between-person object-type condition and within-person surface area control type were associated with performance across training, we used a generalized linear mixed-effects model for the dichotomous outcome of correct/incorrect on each trial during each session. Because the surface area control type varied randomly within person and training was adaptive across control types, children could not advance to a more difficult ratio on one control type and not others. Therefore, correctness (rather than ratio) was the appropriate outcome for this final model.

Linear Mixed-Effects Models. To address our first research question of whether or not children advanced to more difficult ratios, we used ratio on each trial as our outcome variable (recall that a smaller ratio means better performance; see Table 2 for model comparison results). We first ran an intercept-only model predicting ratio, with intercept included as both a fixed and random effect (Model 1). This model allows for children's average ratio at the intercept to be significantly different than 0 (which it had to be because all children started training at a 4:1 ratio) and also allows for variability in the intercept across children. We then added a fixed effect of trial and session (Model 2), which tested for linear effects in ratio slopes across sessions and across trials within sessions. We next added an interaction between trial and session to the model, which, if statistically significant, would suggest that the linear trial slope varied across sessions (Model 3). In other words, it allowed for the possibility that children's improvement across trials within sessions was different across sessions. For example, children may have improved their ratio across trials more in earlier sessions compared with later sessions. We added

TABLE 2
Results and model comparisons of ratio predicted from trial, session, and interactions

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
AIC	20,741	20,715	20,675	20,665	20,178
BIC	20,761	20,749	20,717	20,721	20,268
Log Likelihood	−10,367	−10,352	−10,332	−10,325	−10,076
Deviance	20,735	20,705	20,663	20,649	20,152
Model Comparisons ($\Delta\ \chi^2$)					
Compared to Model 2	(2) = 29.932**				
Compared to Model 3	(3) = 71.588**	(1) = 41.656**			
Compared to Model 4	(5) = 85.466**	(3) = 55.534**	(2) = 13.878**		
Compared to Model 5	(10) = 582.39**	(8) = 552.45**	(7) = 510.8**	(5) = 496.92**	

Note. Models were compared via likelihood ratio tests. ** $p < .01$. Model 1 = intercept-only model, with intercept included as both a fixed and random effect. Model 2 = Model 1 with additional fixed effects added for linear trial and session slopes. Model 3 = Model 2 with an additional interaction term added for the interaction between linear trial and session slopes. Model 4 = Model 3 with additional fixed effects added for quadratic trial and session slopes. Model 5 = Model 4 with additional random effects added for the linear trial and session slopes. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

quadratic fixed effects for trial and session to the next model (Model 4) to test for quadratic change across trials within sessions and for quadratic change in ratio across sessions. The rationale for testing if these effects improved model fit was to allow for the possibility that children would show steep linear change at the beginning of training that would not be fully maintained across training, consistent with the findings of DeWind and Brannon (2012). Also, due to the nature of the adaptive training, children had to complete at least three trials correctly at a particular ratio before moving to a more difficult ratio, which involved inherent nonlinearity in the training design. Finally, we added random effects for the linear trial and session slopes (Model 5) to model interindividual variability in change. Further addition of random effects resulted in issues with convergence, likely due to overparametization of the model. After comparing a series of models (see Table 2) via likelihood ratio tests, the best-fitting model was Model 5.

The parameter estimates for our best-fitting model of change are shown in Table 3. There was a significant intercept fixed effect, indicating that children's overall performance on the first trial of the first session was significantly different from 0 (which it had to be because of the training design). The fixed linear and quadratic effects for both trial and session were statistically significant, and the interaction between linear terms for trial and session was also statistically significant (see Appendix for figures of raw data).

The model-implied interaction between session and trial is plotted is Figure 2. From Figure 2, we see that children improved in ratio across trials within sessions, but they improved more in earlier sessions.

To address if numerically irrelevant stimulus features moderated the effects of training, we used whether or not children were correct on each trial as the primary outcome, and because this was a dichotomous variable (correct or incorrect), we used a logit link function to model this outcome for each trial in a generalized linear mixed-effect model. The estimates obtained from these models, therefore, can be interpreted as: Holding everything else in the model constant, a one-unit change in the predictor of interest has a model-implied change in the logit of the value of the corresponding coefficient. The logit can then be transformed into a probability for ease of interpretation using the following formula:

TABLE 3
Final model-derived parameters for Model 5 with ratio as outcome

Factors	*Estimate*	*95% CI*
Fixed Effects		
Intercept	3.08***	[2.76, 3.40]
Session	−0.11***	[–0.16, –0.06]
Trial	−0.03***	[–0.04, –0.02]
Session × Trial	0.004***	[0.002, 0.005]
Session^2	0.01***	[–0.003, 0.01]
Trial^2	0.001***	[0.0001, 0.001]
SD of Random Effects		
Intercept	0.86	
Session	0.11	
Trial	0.01	

****p* < .001. Parameter estimates are unstandardized.

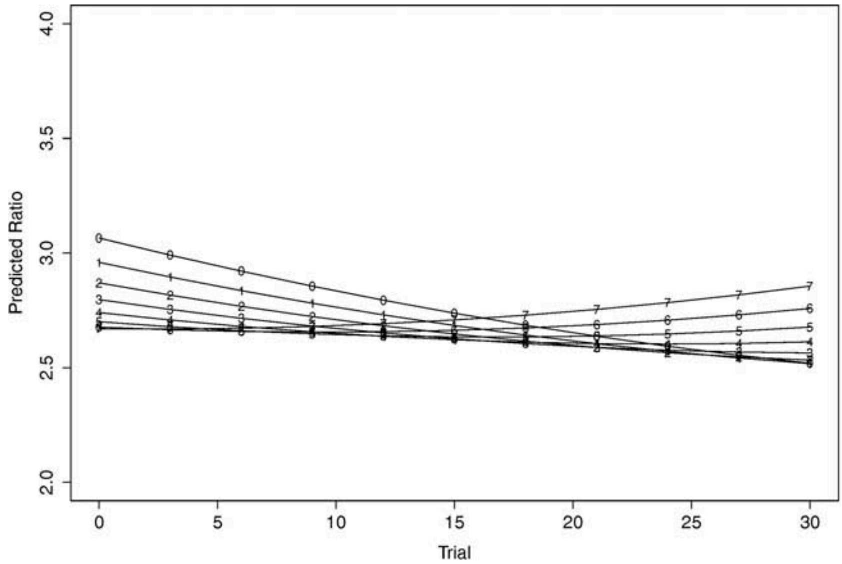

FIGURE 2 The model-implied linear trends for session and trial and the interaction between session and trial on ratio in Experiment 1 (Model 5, Table 3).

TABLE 4
Results and model comparisons of log odds (Logit) of correct responses predicted from trial, session, ratio, number of trials, stimuli, surface area controls, and interactions

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
AIC	9,025	9,028	8,991	8,994	8,976
BIC	9,039	9,056	9,032	9,057	9,060
Log Likelihood	−4,510	−4,510	−4,489	−4,488	−4,476
Deviance	9,021	9,020	8,979	8,976	8,953
Model Comparisons (Δ χ^2)					
Compared to Model 2	(2) = 0.734				
Compared to Model 3		(2) = 41.95**			
Compared to Model 4		(5) = 43.87**	(3) = 1.92		
Compared to Model 5		(8) = 67.63**	(6) = 25.68**	(3) = 23.76**	

Note. Models were compared via likelihood ratio tests. ** $p < .01$. Model 1 = intercept-only model, with intercept included as both a fixed and random effect. Model 2 = Model 1 with additional fixed effects added for linear trial and session slopes. Model 3 = Model 2 with additional fixed effects added for ratio and total number of trials. Model 4 = Model 3 with additional fixed effects added for object type (between-subjects variable) and the interaction between object type and both linear trial and session slopes. Model 5 = Model 4 without object type as a fixed effect and with additional fixed effects for surface area control types as well as interactions between surface area control types and both linear trial and session slopes. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

Probability = exp(logit)/(1+exp(logit)). We again used a model-building approach to test this question (see Table 4). We first compared an intercept-only model (with intercept as fixed and random; Model 1) to a model with fixed effects for trial and session (Model 2).

Although the addition of linear slopes for trial and session did not statistically improve model fit, we proceeded with Model 2 because we were primarily interested in the interaction between the session and trial slopes and numerically irrelevant stimulus features. We first added our control variables of ratio and total number of trials (because children completed different numbers of trials across training) to Model 2 (Model 3). This model was a significantly better fit compared with Model 2. We next added stimuli (object type) to Model 3 as a main effect and also added an interaction between stimuli and both session and trial linear slopes. This model (Model 4) was not a significant improvement in fit compared with Model 3, and neither the main effect of stimuli nor the interactions between stimuli and session and trial were statistically significant (all $p > .05$). This finding means that we do not have statistical evidence that the stimuli object type that children compared affected either their initial performance or their change in performance across sessions and trials. Therefore, we did not retain object type as a factor in our final model. In Model 5, we added surface area control types to Model 3 as main effects and as interaction terms with session and trial linear slopes. Surface area control types were dummy-coded with *mean area equal* trials as the comparison group. The preferred model according to likelihood ratio tests was Model 5 (see Table 4), which included fixed and random effects for intercept, fixed effects for linear trial and session, surface area control type, and interactions between surface area control type and our two linear slope parameters.

Parameter estimates for the best-fitting model are presented in Table 5. We found a main effect for the linear session parameter, but not trial. We found main effects for surface area control type, such that at the intercept (representing initial performance), children were less likely to answer correctly on trials in which surface area was held constant (*total area equal* trials) compared with *mean area equal* trials. Children were also statistically significantly less likely to answer correctly on *inverse* trials compared with *mean area equal* trials. The interactions between surface area control type and session were statistically significant (see

TABLE 5
Final model-derived parameters for Model 5 with correct/incorrect as outcome

Factors	*Estimate*	*95% CI*
Fixed Effects		
Intercept	−1.39	[−3.30, 0.52]
Session	−0.06**	[–0.10, –0.02]
Trial	−0.002	[–0.01, 0.01]
Ratio	0.17**	[0.12, 0.23]
Number of Trials	0.01*	[0.001, 0.02]
Total Area Equal Control	−0.35*	[–0.67, –0.04]
Inverse Control	−0.65**	[–0.95, –0.34]
Total Area Equal Control × Session	0.06*	[0.01, 0.12]
Inverse Control × Session	0.10**	[0.05, 0.16]
Total Area Equal Control × Trial	−0.001	[–0.01, 0.01]
Inverse Control × Trial	0.01	[–0.01, 0.02]
SD of Random Effect	0.58	

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

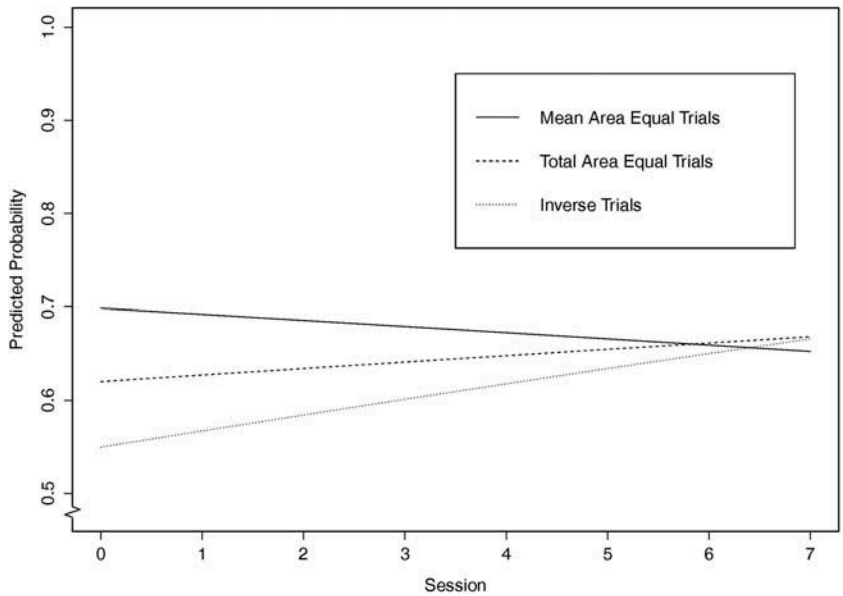

FIGURE 3 The model-implied interaction between surface area control type and session on the probability of answering correctly on each trial in Experiment 1 (Model 5, Table 5).

Table 5 for full model results). As predicted, children significantly improved their performance on *inverse* trials across training sessions compared with changes in *mean area equal* trials. Children also significantly improved their performance on *total area equal* trials across training sessions more than they improved on *mean area equal* trials. Figure 3 displays model-implied interaction data (note that a random effect of 0 represents an individual at the modal value in the plotting of these results).

Discussion

We tested the effects of an adaptive computerized ANS acuity-training paradigm with young children from low-income homes and extended previous research to see if training could be beneficial for children from low-income homes and to determine if the effects of ANS acuity training depend on non-numerical stimulus features. Several important findings emerged. First, we found that children improved their ratio achieved across training designed to explicitly target the nonsymbolic ANS. Children progressed to more difficult ratios within and across sessions; however, children improved more in early training sessions than in later sessions. Secondly, we assessed whether or not children's probability of responding correctly across trials and sessions depended on object type (between-participant variable) or surface area control type (within-participant variable). The specific types of objects children viewed when completing training did not significantly impact their probability of responding correctly, as neither the fixed effect of stimuli nor the interaction between stimuli and slopes (session and trial) were significant. However, we found that the probability of answering correctly across training, controlling for ratio and number of trials, was influenced by the surface area control types. Specifically, children increased their probability of responding correctly across sessions on *inverse* trials and on *total area equal* trials significantly more than on *mean area equal* trials.

We found that children progressed to more difficult ratios across ANS acuity training. Although children significantly decreased their ratio achieved (i.e., improved their performance) across training, their average performance stayed between a 4:1 ratio and a 2:1 ratio, which was generally higher than the 3:2 and 4:3 ratio thresholds reported for preschoolers in Halberda and Feigenson (2008). This discrepancy in performance could be due to the differences in socioeconomic-status backgrounds between the two samples. Children in the Halberda and Feigenson study were from middle- to high-income homes, whereas children in our experiments were from low-income homes. Previous research has shown that children from middle- to high-income homes show significantly better performance than children from low-income homes on tasks designed to rely on the ANS (McNeil et al., 2011). Moreover, when compared with samples of young children who performed above norms on standardized mathematics achievement tests (e.g., Libertus et al., 2011; Libertus, Feigenson, & Halberda, 2013; Mazzocco et al., 2011), children from low-income homes, who scored about 1 standard deviation below norms, showed weaker associations between their ANS acuity and mathematics performance that were reduced to nonsignificance once controlling for inhibitory control (Fuhs & McNeil, 2013). Thus, even modest improvements in ratio could be meaningful for children who are at risk for mathematics difficulties.

We were specifically interested in testing if and how specific non-numerical stimulus features affected children's performance across training. We did not find a significant effect of the specific types of objects used in the arrays. Although solid-colored black dots may be less attention-grabbing and less distracting than colorful, child-friendly objects such as cows, pigs, shoes, or balls, children did not benefit more from training that included solid-colored black dots than from training that included colorful objects. It may be that the benefits and disadvantages of the two stimuli types canceled each other out. Specifically, children who trained with perceptually rich, concrete objects may have been more interested and engaged in the task, but at the same time, they may have had more difficulties focusing on the numerosity of the arrays in the face of the competing non-numerical features of the actual objects to be compared.

When we compared children's probability of answering correctly across different surface area control types, we found that children improved significantly more on trials in which numerosity was inversely related to surface area relative to trials in which surface area was positively correlated or congruent with numerosity, similar to training results recently reported in adults (DeWind & Brannon, 2012). Children also improved significantly more on trials in which total surface area was held constant across object sets compared with trials in which surface was positively correlated with numerosity. Although we did not explicitly compare *inverse* slopes to *total area equal* slopes because *mean area equal* trials were the comparison group, Figure 3 suggests that the steepest positive slope was seen for *inverse* trials compared with both *total area equal* trials and *mean area equal* trials. *Mean area equal* slopes appeared to decrease relative to *total area equal* and *inverse* slopes, which could suggest that children's initially high performance on the *mean area equal* trials could have been due to a heavy reliance on surface area cues that were congruent with numerosity. In other words, children's initial bias toward surface area may have yielded artificially high performance in the *mean area equal* trials at the beginning of the training. After some training, however, children may have learned to shift their focus away from surface area and toward numerosity, thus leading to decreased performance on these trials and increased performance on the other two trial types that require a focus on numerosity for a correct response. An overall interpretation of the interaction between training and surface area control type could be that training improves children's access to their existing ANS by teaching them to inhibit conflicting or irrelevant stimulus features including total surface area of the object sets and mean surface area of the object sets.

There were at least two limitations of the present experiment that needed to be addressed in a second experiment. First, we could not rule out the possibility that children adopted a strategy that would allow them to succeed on the *total area equal* and *inverse* trials without actually improving their ANS acuity. Specifically, in both the *total area equal* trials and the *inverse* trials, the object set with more objects also had a smaller mean object size, whereas the *mean area equal* trials equated the average object size. Thus, it is possible that children learned to pick the side with the smaller average dot size—a strategy that would have been beneficial on *total area equal* and *inverse* trials but would have led to decreased performance on *mean area equal* trials.

Secondly, we could not definitively say that the ANS training is what improved children's attention to numerosity because we did not include a group of children who did not receive any training to rule out the possibility that maturation or other experiences in the preschool classroom contributed to the improvements on the *total area equal* and *inverse* trials over time. Education research has shown that preschool children are engaged in mathematics-related activities on average for only 6% of their day (Frank Porter Graham [FPG] Child Development Institute, 2005), suggesting it is unlikely that this very small fraction of time spent in mathematics-related activities could account for the ratio improvements in this study. However, it was still important to rule out alternative explanations for training effects. Thus, in the second experiment, we randomly assigned children to either an experimental ANS training group or a control group to rule out maturation and other preschool experiences as confounds. Importantly, we also added a pretest and posttest measure of children's ANS acuity that was separate from the training trials to examine if ANS acuity training led to significantly better performance on an ANS acuity task, particularly on the *inverse* trials, when compared with children in the control condition. We included additional ANS dot-size trials in this assessment in which the side with more dots had a larger average dot size equal to the ratio between the two arrays to test if children had adopted the strategy of picking the set with the smaller average dot size.

EXPERIMENT 2

Method

Participants. Approval for this project was obtained from a university Human Subjects Review Board as well as from the Head Start program from which the participants were drawn. Fifty-four children from Head Start programs were assigned to either an adaptive ANS acuity-training intervention or a shared book-reading (control) condition. Seven children were excluded due to experimenter error (n = 3), refusal to complete the activities (n = 2), or being absent for more than three sessions (n = 2). Thus, the final sample included 47 children (M_{age} = 4;9, SD = 6.69 months; 20 girls, 27 boys; 15% White, 66% African American or Black, 15% Hispanic, and 4% Other).

Design. The design was a randomized experiment in which children were randomly assigned to an ANS training intervention or to a control intervention in which they participated in shared book reading (Justice & Ezell, 2002).

ANS Acuity Measure. Children were shown two separate dot arrays presented on a white sheet of paper and were asked to pick, without counting, which side had "more dots." The measure consisted of five different ratios (2:1, 3:2, 4:3, 5:4, and 6:5), with 7 trials at each ratio, for a total of 35 trials. The ratio and trial type were randomly selected without replacement for each of the 35 trials when the acuity measure was constructed. There were four types of trials within the ANS acuity test, three of which matched the *mean area equal, total area equal*, and *inverse* trials from Experiment 1. For each ratio, children saw 2 of each of these trials. In addition to these trials, a fourth trial type was included to investigate whether training was leading children to simply pick the side with the smaller average dot size. In these dot-size trials, the side with more dots had a larger average dot size equal to the ratio between the two arrays (e.g., the side with twice as many dots had an average dot size that was twice that of the array with fewer dots). Each of the five ratios included one dot-size trial. Overall performance on the ANS acuity task was measured as performance on the mean area, total area, and inverse trials only (i.e., number correct out of 30 trials). Performance on the 5 dot-size trials was analyzed separately.

Experimental ANS Training Task. The experimental ANS training task in Experiment 2 was made to closely mirror that in Experiment 1. However, because the Head Start director suggested that it would be more beneficial for children to work one-on-one with a tutor rather than at a computer, we modified our adaptive computer program from Experiment 1, so it could be administered via paper in one-on-one training sessions with a tutor. To facilitate the one-on-one paper version of the task, a few modifications were made. Children were presented with two arrays of dots on a sheet of paper, each enclosed in a rectangular box, and were asked to choose which side had "more dots." The images used in the training were generated from the computer program used during Experiment 1. To maintain the adaptive nature of the training, the trials were set up in blocks of nine trials at a given ratio, and performance on a given block determined the difficulty of the next block. There were three trials of each of the different original trial types within each block of nine (*mean area equal* trials, *total area equal* trials, and *inverse* trials).

On each trial, children were verbally prompted by the tutor to get ready (e.g., "OK. Ready for the next one?"). Following the prompt, the tutor held up the paper a short distance away from the child and said, "Which side has *more* dots?" Children were given an unlimited amount of time to respond, but they were discouraged from counting. Children indicated their responses by pointing to the side with more dots. Following children's responses, feedback was provided. The feedback was made to mirror what children heard during Experiment 1 (e.g., "You got it; way to go!" for correct responses, or "Whoops! You missed that one; try again" for incorrect responses).

All children began training at the 4:1 ratio block, and the training was adaptive across sessions, such that children started the next session on the ratio block that was 2 ratios easier than the ratio block where they had left off in the previous session. If children got seven or more trials (out of nine) correct, then training proceeded to the next most difficult ratio. If a child answered five or fewer trials correctly, then they were moved back to the closest easier ratio block. For example, if a child answered five of nine trials in the 4:3 ratio block incorrectly, then he or she moved to the 3:2 ratio block. If a child answered exactly six out of the nine trials correctly, they repeated a new set of nine trials at the same ratio. There were 7 different ratio possibilities children could complete: 4:1, 2:1, 3:2, 4:3, 5:4, 6:5, and 7:6. Experiment 1 had 10 possible ratios, but the 3 most difficult ratios were rarely, if ever, achieved. For each ratio, the specific numbers of dots in the two boxes were randomly selected from among 12 possibilities.

To encourage engagement with the task, children were introduced to the task as a game. For each correct response, they received a “token” (a colorful foam square), but for each incorrect response, the experimenter earned a token. At the end of each set (i.e., after nine trials at a given ratio), the person with more tokens would receive an extra sticker. This was done because children found it hard to stay engaged with the initial paper version of the task where the experimenter only provided feedback. Once the task was introduced as a game, however, children were motivated to succeed and often remarked about how fun the task was.

Procedure. Children typically worked with the tutors 2 days a week (e.g., Monday and Wednesday or Tuesday and Thursday) during the course of a 3-week time period. Any absences were made up; however, as mentioned, we excluded two children who were absent more than three times during the intervention period. In the first session, children completed the ANS acuity test, followed by a ~7-min ANS acuity-training session or shared book-reading session (depending on condition). Next, children completed two additional ~12-min ANS training or shared book-reading sessions. Finally, children completed their last ~7-min ANS acuity or shared book-reading session followed by the ANS acuity test. The overall amount of training was less than the amount used in Experiment 1 because children in Experiment 1 made most of their improvements early in the training. All children completed training or book reading one-on-one with a tutor in a quiet area of their school.

Results

Random assignment worked as intended and resulted in well-matched groups. Children in the two intervention conditions had similar numbers of girls and boys (ANS acuity, 14 boys and 10 girls; control, 13 boys and 10 girls) and a similar distribution of race/ethnicity (ANS acuity, 17% White, 62% African American or Black, 17% Hispanic, and 4% Other; control, 13% White, 70% African American or Black, 13% Hispanic, and 4% Other). The two groups also had similar scores on the ANS acuity pretest (ANS acuity, M = 20.00, SD = 5.40; control, M = 20.12, SD = 5.14), and this finding held for all trial types, including the *mean area equal* trials (ANS acuity, M = 7.59, SD = 1.89; control, M = 7.46, SD = 1.75), the *total area equal* trials (ANS acuity, M = 6.48, SD = 2.26; control, M = 6.88, SD = 1.86), and the *inverse* trials (ANS acuity, M = 5.93, SD = 2.50; control, M = 5.77, SD = 2.54). The number of trials each child completed, the number of correct trials, and the mean ratio of the trials within each session are displayed in Table 6. As with Experiment 1, children’s average ratios achieved were not as small as ratio thresholds reported from studies with children from middle- and high-income families. However,

TABLE 6
Experiment 2 descriptive statistics within sessions for experimental group

Session		N	*Minimum*	*Maximum*	*M*	*SD*
1	Total Trials	24	7	36	25.42	8.43
	Total Correct Trials	24	2	35	21.75	9.05
	Average Ratio Achieved	24	2.21	4.00	2.80	0.66
2	Total Trials	24	18	63	40.88	10.27
	Total Correct Trials	24	13	56	32.71	10.49
	Average Ratio Achieved	24	1.29	4.00	2.00	0.70
3	Total Trials	24	18	63	39.38	10.88
	Total Correct Trials	24	10	49	29.46	9.46
	Average Ratio Achieved	24	1.17	4.00	1.71	0.68
4	Total Trials	24	7	63	22.42	11.68
	Total Correct Trials	24	3	48	16.42	9.48
	Average Ratio Achieved	24	1.18	4.00	1.88	1.04

Note. For average ratio achieved, smaller numbers indicate *better* performance.

children achieved a better average ratio (about 2:1) in Experiment 2 than they did in Experiment 1 (ratio ranging from 4:1 to 2:1).

Analyses of Covariance. We performed an analysis of covariance (ANCOVA) to examine if the children who had received the ANS acuity training exhibited better ANS acuity—particularly on the inverse trials—than did the children who had received the shared book-reading control intervention. We then performed an additional ANCOVA on the dot-size trials to test if the ANS acuity training led children to adopt a strategy of picking the side with smaller dots.

First, to examine if the ANS intervention affected performance on the different ANS trial types differently, we performed a mixed-factor ANCOVA with intervention condition (ANS acuity or control) as the between-subjects factor, ANS trial type (*mean area equal, total area equal, inverse*) as the within-subjects factor, performance on the ANS acuity task at pretest as the covariate, and number correct (out of 10) at posttest as the outcome. The effect of pretest ANS acuity was significant, $F(1, 44) = 26.28$, $p < .001$, partial eta squared = .37, with higher pretest scores associated with higher posttest scores. The main effects of condition and trial type were not significant, $F(1, 44) = 1.88$, $p = .18$, partial eta = .04, and $F(2, 88) = 1.68$, $p = .19$, partial eta = .04, nor was the interaction of pretest ANS acuity and trial type, $F(2, 88) = 0.42$, $p = .66$, partial eta = .01. There was, however, evidence of the predicted interaction between condition and trial type, $F(2, 88) = 3.00$, $p = .05$, partial eta = .06. As predicted, simple-effects tests showed a simple main effect of intervention condition on *inverse* trials, $F(1, 111.66) = 5.96$, $p = .02$, but not on *mean area equal* trials, $F(1, 111.66) = 0.21$, $p = .65$, or *total area equal* trials, $F(1, 111.66) = 1.04$, $p = .31$. Figure 4 displays the adjusted means on the ANS acuity task at posttest as a function of condition. As can be seen in the figure, children who received the ANS acuity intervention answered more inverse trials correctly ($M = 6.61$, $SD = 1.96$) than did children who received the control intervention ($M = 5.45$, $SD = 1.97$, $d = 0.59$).

Importantly, we found no evidence that this superior performance on the inverse trials was due to the development of a strategy to select the side with smaller dots. Recall that we included five supplementary trials in this experiment in which the average dot size of the side with more

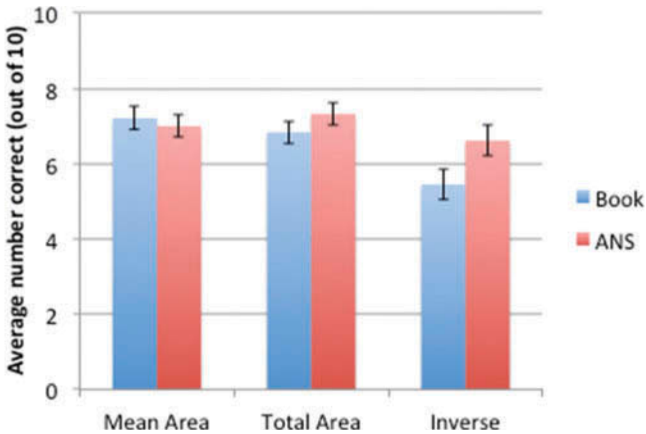

FIGURE 4 Average number correct on the ANS acuity posttest as a function of condition and surface area control type in Experiment 2. *Note*. Values control for approximate number system (ANS) acuity pretest scores.

dots also was much greater than the side with fewer dots. An ANCOVA with intervention condition (ANS acuity or control) as the independent variable, pretest ANS acuity as the covariate, and number correct on the five supplementary trials as the outcome showed no evidence of a main effect of condition, $F(1, 46) = 0.057$, $p = .81$, partial eta squared = .001. Contrary to what would be expected if children had adopted the strategy of choosing the side with the smaller average dot size, children performed quite well on the supplementary trials, and those who received the ANS acuity intervention correctly solved about the same number of problems ($M = 4.40$, $SD = 1.03$) as children in the control intervention ($M = 4.33$, $SD = 1.00$).

Discussion

We addressed the major limitation of Experiment 1 by conducting a randomized experiment to see if the ANS acuity training led to better performance on *inverse* ANS trials at posttest when compared to a shared book-reading control intervention. Controlling for performance on the ANS task at pretest, we found an interaction between condition and ANS task trial type, and simple main-effect tests showed that children who had received the ANS acuity training performed significantly better on *inverse* trials compared with children who had received the shared book-reading intervention. Importantly, we found that the benefits of the ANS acuity training on inverse trials could not be explained by children adopting the strategy to pick the set with smaller dots.

GENERAL DISCUSSION

Much of the research on ANS acuity development and training has focused on children from middle- to high-income homes rather than on children who are at increased risk for mathematical difficulties due to living in a low-income household. We know from previous research that children growing up in low-income homes have less access to mathematics materials in their homes and less access to high-quality early childhood experiences from which to learn early mathematics concepts (e.g., Bradley & Corwyn, 2002; Ramani & Siegler, 2008; Votruba-Drzal, 2003). Therefore, it is important to understand how ANS acuity training affects children in this disadvantaged group because children living in poverty may benefit most from ANS training if it can benefit children's ANS acuity.

Both experiments yielded evidence that children benefitted from ANS acuity training, but these benefits were generally limited to *inverse* trials rather than across-the-board improvements on different trial types. In the first experiment, we found that children improved significantly more on *inverse* and *total area equal* trials compared with *mean area equal* trials across sessions. In the second experiment, we found that children who had received the ANS acuity training made more gains from pretest to posttest than did children who had received the control intervention only on the *inverse* trials. Although children in the first experiment improved significantly more on *total area equal* trials compared with *mean area equal* trials, an examination of Figure 2 suggests that the positive slope for *inverse* trials was steeper than that of the *total area equal* trials. Thus, taken together, the two experiments suggest that improvements in ANS acuity training for children from low-income homes may be driven by improvements in their ability to inhibit attention to conflicting surface area cues.

Children may initially show preference for using surface area cues over numerosity cues, particularly if these cues conflict with numerosity cues, but over time, they may become increasingly sensitive to numerosity (e.g., Clearfield & Mix, 1999, 2001; Gebuis & Reynvoet, 2011; Rousselle, Palmers, & Noël, 2004; Soltész et al., 2010). If sensitivity to numerosity cues in the environment is aided by exposure to numerical activities, children from low-income homes may benefit from activities that aid in the process of attending to numerosity cues. The results of the current experiments suggest that training can speed up this process such that increased exposure to ANS tasks with feedback allows children to become increasingly sensitive to numerosity over surface area cues. This interpretation would explain why children initially showed much better performance on *mean area equal* trials in Experiment 1 but then showed a decrease in performance on those trials relative to the other trial types across sessions. It would also explain why, across both experiments, children initially showed the poorest performance on trials in which there was a conflict between numerosity and total and mean surface area of the object sets. This interpretation would additionally explain why condition differences in Experiment 2 were evident on *inverse* trials specifically but not for other trial types. For children from low-income homes who may not have as many experiences with mathematics in their homes prior to schooling, training could potentially improve children's ability to attend to numerosity cues in their environment.

It is important to note that the effects of training on the *inverse* trials of the ANS acuity posttest in Experiment 2 were small, and questions remain as to whether or not these results are practically meaningful for children's early learning. The results of the current study are related to other training studies in which researchers have attempted to directly train children to attend to a specific dimension while ignoring irrelevant stimuli in a domain-neutral context. Interestingly, training studies have previously shown that exposure to tasks in which children are required to show inhibitory control, or focusing on relevant dimensions in the face of irrelevant or distracting information, yield significant improvement in children's inhibitory control, even on non-trained tasks (e.g., Dowsett & Livesey, 2000; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005; Tominey & McClelland, 2011). For example, Tominey and McClelland (2011) found that children who participated in games in which they had to inhibit a prepotent response (e.g., not play an instrument when the experimenter waved her conducting baton but playing when she did not wave her conducting baton) improved their inhibitory control abilities on a nontrained task at the conclusion of training.

It could be argued then that the ANS acuity-training paradigm, focused on attention to numerically relevant information and ignoring irrelevant surface area cues, actually works by increasing children's inhibitory control, specifically in the domain of mathematics. It has been well established that children's inhibitory control skills are related to their symbolic mathematics skills (e.g., Blair & Razza, 2007; Bull, Espy, & Wiebe, 2008; Bull & Scerif, 2001; Clark, Pritchard, & Woodward, 2010; Espy et al., 2004; Fuhs, Nesbitt, Farran, & Dong, 2014; St Clair-Thompson & Gathercole, 2006; van der Sluis, de Jong, & van der Leij, 2007). Researchers have also shown that children's inhibitory control is associated with children's ANS acuity (Clayton & Gilmore, 2015; Fuhs & McNeil, 2013; Gilmore et al., 2013). Inhibitory control improvements may then drive improvements on both ANS acuity tasks and symbolic mathematics tasks.

These findings have implications for competing explanations of individual differences in ANS acuity. Specifically, the results support the competing-processes account of ANS acuity (Clayton & Gilmore, 2015; Gebuis & Reynvoet, 2011). According to this account, children's performance on ANS acuity tasks is a result of numerical and visual representations as well as inhibitory control processes required to lessen distraction to competing non-numerical stimulus features. We found in Experiment 1 that children's performance on *mean area equal* trials appeared to decrease while performance on *inverse* trials and *total area equal* trials increased. In Experiment 2, ANS acuity training translated to significant gains on *inverse* trials but not *mean area equal* trials. These results suggest that this increased ability to inhibit attention to surface area cues in favor of attending to numerosity drove the intervention effects. To further explore the competing-processes account, future work is needed to explicitly assess whether or not inhibitory control may explain improvement in young children's ANS acuity and also if inhibitory control interventions could improve children's ANS acuity performance and produce larger domain-general effects on early learning.

Limitations and Future Directions

A few limitations should be noted. First, nearly all of the across-session ratio improvements in Experiment 1 occurred in the first few sessions. These results are similar to those of other attempts to train ANS acuity in which most of the improvements in ratio were seen in the first few sessions (e.g., DeWind & Brannon, 2012), and they raise important questions about what children were actually learning during training. One possibility is that children's early improvements in ratio were simply the result of gaining familiarity with the vocabulary and behavioral responses necessary to perform the task (e.g., understanding what "sides" mean, knowing when to focus attention on the computer screen versus the keyboard, figuring out which button to press to indicate the desired side, etc.) and were not indicative of true changes in children's underlying ANS ratio thresholds or even in children's ability to access their ANS. However, if all of the observed improvements were due to children gaining familiarity with the task, then we might have expected to see more similar gains across the three surface area control trial types, rather than seeing most gains on the inverse trials.

Another possibility is that the design of the adaptive computerized training program in Experiment 1 was not ideal for continued improvement in ratio across sessions. We designed the computer training to be adaptive not only within sessions, but also across sessions. In other words, if a child ended a session at a 2:1 ratio, then that child would begin the subsequent session at a 2:1 ratio. This aspect of the training regimen may have been less than ideal, as recent research has suggested that beginning a session at a more difficult ratio may affect children's confidence and subsequent performance during training (Odic, Hock, & Halberda, 2014). We modified the between-session difficulty in the second experiment such that children began each new training session at two ratio levels below their performance at the end of the previous session. Thus, children did not begin each session at a too-difficult ratio in Experiment 2. Although we did not explicitly compare the ratios achieved across studies, an examination of Tables 1 and 6 suggests that children in Experiment 2 potentially benefitted from this modification, as they reached lower ratios in training. However, other explanations for the difference across experiments exist, including that differences were based on the use of a computer versus paper versions of the task or the difference in number of sessions across studies, to name a few. The use of an adaptive program has been shown to be an ideal training regimen for mathematics (e.g., Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009), and thus, future research should investigate the effect of ANS acuity training that is adaptive within each session but not across sessions to determine if this procedure results in increased ratio improvements beyond what was found in the current experiments.

Still another explanation for the seeming lack of ratio improvements after the first few sessions could be that the overall ratio improvements were washed out by children's improvements in attention to numerosity. Recall that the surface area control types varied within ratio, and children's bias to respond based on surface area was reduced across the sessions, yielding improvements on *inverse* and *total area equal* trials but decrements on *mean area equal trials*. If children improve on incongruent trials within each ratio but get worse on congruent trials, then they would be expected to continue to hover around the same overall ratio achieved according to the adaptive criteria of the training program. Future studies using alternative designs to try to disentangle ratio and surface area bias effects will help provide information about the nature of the ratio effects.

Another important direction for future research is to assess how children's developing understanding of the term "more" may be affected by ANS acuity training. It is possible that the benefits of ANS acuity training in the current experiments were due, at least in part, to children learning to interpret "more" in the ANS acuity task to mean that the set of dots has more in terms of numerosity, regardless of other features of the set (Negen & Sarnecka, 2015). All trials indicated that children were to focus on numerosity by including only discrete quantities in which children were asked to decide which array had more of a particular count noun, signaling numerosity (as opposed to mass nouns, e.g., water). The recent work of Odic and colleagues (Odic, Pietroski, Hunter, Lidz, & Halberda, 2013) suggests that children may have an immediate acquisition of the word "more" such that they can understand the word "more" to refer to both count and mass nouns at around 3 years of age, younger than the average age of the current sample. However, their work did not include a sample of children from low-income homes, and it will be important in future work to further explore linguistic cues as they relate to children's ANS acuity.

Finally, based on these data alone, we do not know whether or not ANS training paradigms will lead to significant improvement in children's general mathematics achievement, which is an important question to address in future research among children who are at risk for mathematics difficulties. Previous studies that have attempted to ANS acuity training using nonsymbolic numerical comparison alone have produced mixed findings with respect to whether such training improves symbolic mathematics. For example, Park and Brannon (2014) found that training with nonsymbolic numerical comparison alone did not lead to improvements in nontrained symbolic mathematics abilities in adults, whereas nonsymbolic approximate addition training did. In contrast, Hyde and colleagues (2014) found that brief nonsymbolic numerical comparison training and nonsymbolic approximate addition training both led to improvements on a non-trained symbolic mathematics task for children. Thus, it is possible that training with nonsymbolic numerical comparison alone is beneficial for mathematics achievement only up to a certain age or knowledge level. However, neither of the previous studies was conducted with young children from low-income homes, so it is difficult to generalize their results to young children who may be at risk for mathematics difficulties. Given conflicting findings in the literature across studies of children and adults of varying ages and socioeconomic-status backgrounds and with varying experience in mathematics, a developmental perspective will be necessary to delineate the potential benefits of different types of ANS acuity training. It will be important in future studies to conduct long-term follow-ups of these training programs as it remains possible that attention to numerosity amid distractions is a critical component of early development of ANS acuity but may play less of a role as children age and gain more experience with mathematics in their environment.

Conclusion

Our results suggest that one possible process through which children from low-income homes improve their performance on ANS acuity tasks is through an increased ability to disentangle numerosity from other stimulus features, specifically surface area features. Consistent with previous work with children from middle- and high-income homes, we found that ANS acuity training led to improved ratio thresholds. However, in the current experiments, the influence of training on ANS acuity was generally limited to *inverse* trials. More work is needed to determine if improvements in disentangling numerosity from other competing non-numerical features improves children's symbolic mathematics achievement. Overall, these findings provide evidence of the malleability of ANS acuity on certain trials in children who are at risk for poor mathematics outcomes and suggest that one mechanism for improvement is an increasing ability to focus on numerosity in the face of conflicting stimulus features.

ACKNOWLEDGMENTS

We are grateful for the support of teachers, administrators, and parents at the Head Start program from which the participants were drawn.

REFERENCES

Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological Science*, *25*, 2017–2026. doi:10.1177/0956797614547539

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*, 201–221. doi:10.1016/S0010-0277(02)00178-6

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences*, *102*, 14116–14121. doi:10.1073/pnas.0505512102

Bates, D., Maechler, M., & Bolker, B. (2011). Lme4: Linear mixed-effects models using S4 classes (Version 0.999375-42). Retrieved from http://cran.R-project.org/package=lme4

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*, 647–663. doi:10.1111/j.1467-8624.2007.01019.x

Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology*, *114*, 375–388. doi:10.1016/j.jecp.2012.09.015

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–399. doi:10.1146/annurev.psych.53.100901.135233

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*, 205–228. doi:10.1080/87565640801982312

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*, 273–293. doi:10.1207/S15326942DN1903_3

Burns, M. (1996). How to make the most of math manipulatives. *Instructor*, *105*, 45–51.

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

Clayton, S., & Gilmore, C. (2015). Inhibition in dot comparison tasks. *ZDM - The International Journal on Mathematics Education*, *47*, 759–770. doi:10.1007/s11858-014-0655-2

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, *10*, 408–411. doi:10.1111/1467-9280.00177

Clearfield, M. W., & Mix, K. S. (2001). Amount versus number: Infants' use of area and contour length to discriminate small sets. *Journal of Cognition and Development*, *2*, 243–260. doi:10.1207/S15327647JCD0203_1

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks Project. *Journal for Research in Mathematics Education*, *38*, 136–163.

Davison, A. (2005). *Killer game programming in Java*. Sebastopol, CA: O'Reilly.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics* (rev. and updated ed.). New York, NY: Oxford University Press.

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, *103*, 469–479. doi:10.1016/j.jecp.2009.01.010

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience*, *6*, 68. doi:10.3389/fnhum.2012.00068

Dowsett, S. M., & Livesey, D. J. (2000). The development of inhibitory control in preschool children: Effects of 'executive skills' training. *Developmental Psychobiology*, *36*, 161–174. doi:10.1002/(SICI)1098-2302(200003)

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities*, *46*, 166–181. doi:10.1177/0022219411410233

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology*, *26*, 465–486. doi:10.1207/s15326942dn2601_6

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

Frank Porter Graham (FPG) Child Development Institute. (2005). *How is the pre-k day spent? Early development*. Chapel Hill: The University of North Carolina at Chapel Hill, FPG. Retrieved from http://fpg.unc.edu/node/260

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science*, *16*, 136–148. doi:10.1111/desc.12013

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology*, *50*, 1698–1709. doi:10.1037/a0036633

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods*, *43*, 981–986. doi:10.3758/s13428-011-0097-5

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

Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*, 414–466. doi:10.1016/S0010-0285(02)00519-4

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the 'Number Sense': The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*, 1457–1465. doi:10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences, 109*, 11116–11120. doi:10/1073/pnas.1200196109

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

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

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*, 92–107. doi:10.1016/j.cognition.2013.12.007

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review*, *18*, 1222–1229. doi:10.3758/s13423-011-0154-1

Justice, L. M., & Ezell, H. K. (2002). Use of storybook reading to increase print awareness in at-risk children. *American Journal of Speech-Language Pathology*, *11*, 17–29. doi:10.1044/1058-0360(2002/003)

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantages of abstract examples in learning math. *Science*, *320*, 454–455. doi:10.1126/science.1154659

Keller, L., & Libertus, M. (2015). Inhibitory control may not explain the link between approximation and math abilities in kindergarteners from middle-class families. *Frontiers in Developmental Psychology*, *6*, 685. doi:10.3389/fpsyg.2015.00685

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, *1*, 155–178. doi:10.1080/19345740802114533

Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science*, *32*, 366–397. doi:10.1080/03650210701863933

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*, 1292–1300. doi:10.1111/desc.2011.14.issue-6

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences*, *26*, 126–133. doi:10.1016/j.lindif.2013.02.001

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science*, *21*, 873–881. doi:10.1177/0956797610370158

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

McNeil, N. M., Fuhs, M. W., Keultjes, M. C., & Gibson, M. H. (2011). Influences of problem format and SES on preschoolers' understanding of approximate addition. *Cognitive Development*, *26*, 57–71. doi:10.1016/j.cogdev.2010.08.010

McNeil, N. M., Uttal, D. H., Jarvin, L., & Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction*, *19*, 171–184. doi:10.1016/j.learninstruc.2008.03.005

Negen, J., & Sarnecka, B. W. (2015). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *British Journal of Developmental Psychology*, *33*, 92–105. doi:10.1111/bjdp.1207

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS One*, *8*, e67918. doi:10.1371/journal.pone.0067918

Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., & Content, A. (2013). Does math education modify the approximate number system? A comparison of schooled and unschooled adults. *Trends in Neuroscience and Education*, *2*, 13–22. doi:10.1016/j.tine.2013.01.001

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General*, *142*, 255–265. doi:10.1037/a0030825

Odic, D., Pietroski, P., Hunter, T., Lidz, J., & Halberda, J. (2013). Young children's understanding of 'more' and discrimination of number and surface area. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *39*, 451–461. doi:10.1037/a0028874

Office of Head Start. (n.d.). Retrieved from http://www.acf.hhs.gov/programs/ohs

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, *24*, 2013–2019. doi:10.1177/0956797613482944

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

Petersen, L. A., & McNeil, N. M. (2013). Effects of perceptually rich manipulatives on preschoolers' counting performance: Established knowledge counts. *Child Development*, *84*, 1020–1033. doi:10.1111/cdev.12028

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*, 375–394. doi:10.1111/j.1467-8624.2007.01131.x

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development*, *24*, 450–472. doi:10.1016/j.cogdev.2009.09.003

R Development Core Team. (2012). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rousselle, L., Palmers, E., & Noël, M. (2004). Magnitude comparison in preschoolers: What counts? Influence of perceptual variables. *Journal of Experimental Child Psychology*, *87*, 57–84. doi:10.1016/j.jecp.2003.10.005

Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences*, *102*, 14931–14936. doi:10.1073/pnas.0506897102

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Quarterly Journal of Experimental Psychology*, *67*, 271–280. doi:10.1080/17470218.2013.803581

Sloutsky, V. M., Kaminski, J. A., & Heckler, A. F. (2005). The advantage of simple symbols for learning and transfer. *Psychonomic Bulletin & Review*, *12*, 508–513. doi:10.3758/BF03193796

Soltész, F., Szűcs, D., & Szűcs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: A developmental study. *Behavioral and Brain Functions*, *6*, 13–14. doi:10.1186/1744-9081-6-13

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Infants show ratio-dependent number discrimination regardless of set size. *Infancy*, *18*, 927–941. doi:10.1111/infa.12008

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, *59*, 745–759. doi:10.1080/17470210500162854

Stoianov, I., & Zorzi, M. (2012). The emergence of a visual number sense in hierarchical generative models. *Nature Neuroscience*, *15*, 194–196. doi:10.1038/nn.2996

Tominey, S. L., & McClelland, M. M. (2011). Red light, purple light: Findings from a randomized trial using circle time games to improve behavioral self-regulation in preschool. *Early Education & Development*, *22*, 489–519. doi:10.1080/10409289.2011.574258

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*, 427–449. doi:10.1016/j.intell.2006.09.001

Votruba-Drzal, E. (2003). Income changes and cognitive stimulation in young children's home learning environments. *Journal of Marriage and Family*, *65*, 341–355. doi:10.1111/j.1741-3737.2003.00341.x

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483–488. doi:10.1016/j.tics.2003.09.002

Whyte, J. C., & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology*, *44*, 588–596. doi:10.1037/0012-1649.44.2.588

Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education*, *3*, 224–234. doi:10.1111/mbe.2009.3.issue-4

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

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, *8*, 88–101. doi:10.1111/desc.2005.8.issue-1

APPENDIX

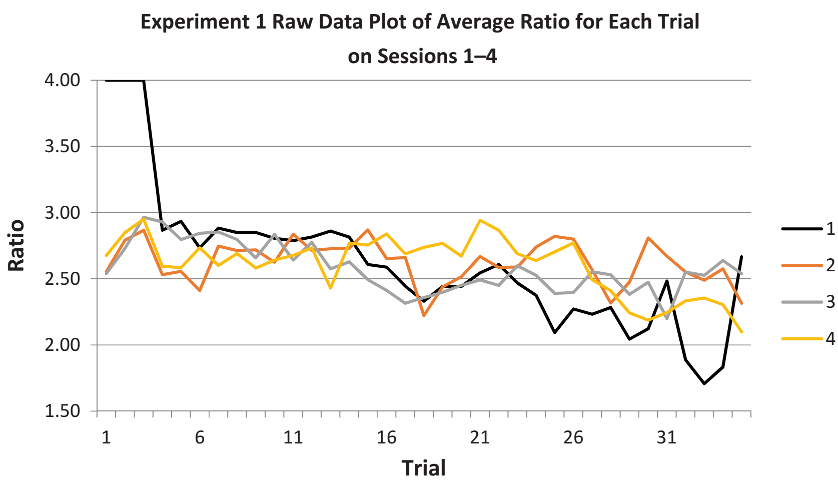

FIGURE 1 Experiment 1 raw data plot of average ratio for each trial on Sessions 1 through 4.

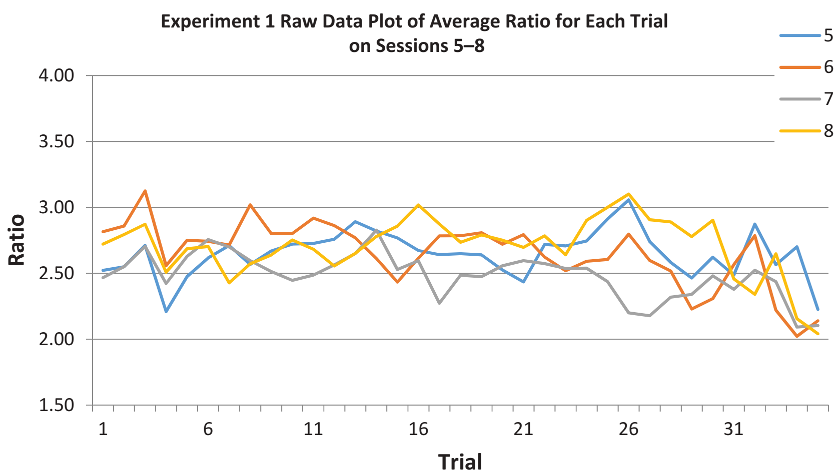

FIGURE 2 Experiment 1 raw data plot of average ratio for each trial on Sessions 5 through 8.

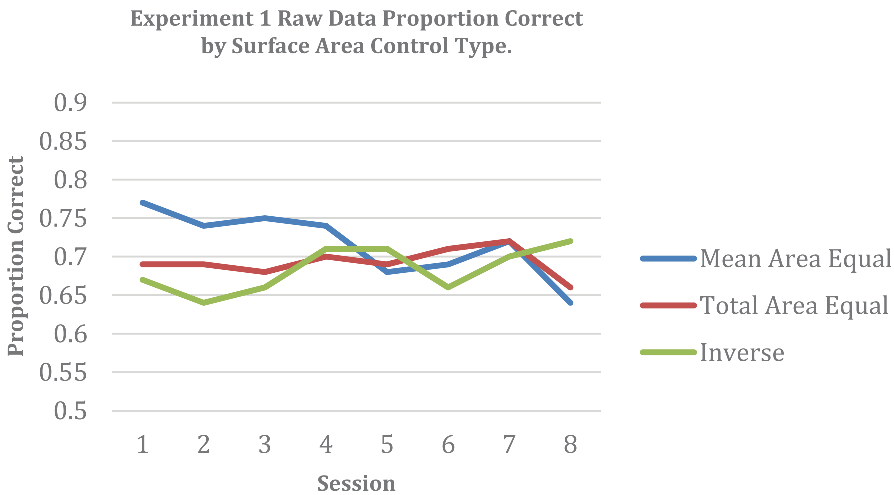

FIGURE 3 Experiment 1 raw data proportion correct by surface area control type.