Flexible Attention to Magnitudes and Early Math

NCT05682729

6/30/2024

Method

Participants

Participants (N = 116) in the current study were children recruited from Head Start preschool programs in an urban Midwest city and had a mean age of 55.6 months (SD = 7.8). Participants were 54.3% Female and 45.7% Male. The sample was 41.4% White, 27.6% Black or African American, 15.5% multi-racial (mostly consisting of White and Black or African American), 4.3% identified as "Other," and 11.2% chose not to respond. Additionally, 8.6% of children were Hispanic in ethnicity. Child-level data on socioeconomic status was not collected because all children attended Head Start centers. Head Start is a federally funded early childhood program in the United States for children in families with an income at or below the poverty line, who are experiencing homelessness, who are in foster care, or who are receiving public benefits such as Temporary Assistance for Needy Families (TANF) or Supplemental Security Income (SSI) (Head Start | *Ohio Department of Education and Workforce*, 2024). Consent forms were signed by parents/caregivers before data were collected, and children gave verbal assent before participating. In the case of absences, if a participant missed one reading session, that session was made up prior to the completion of the post-intervention test. Participants who were absent for more than one reading session were excluded from the analyses. All participants included in the analysis received all three intervention sessions.

Procedure

The study procedures took place in the same order for each child, starting with the FAM assessment and followed by a number line estimation task, the Woodcock-Johnson Number Sense subtest, and the Minnesota Executive Functioning Scale (MEFS). Tasks were administered in a fixed order to allow us to compare task performance across participants without order effects

obscuring the findings of the analysis. Data collection and intervention took place over a four-day period for each child in a quiet area of the classroom to minimize distractions. In the event of an absence during one of the days, the data collection period extended to five days. On day one, the four assessments were administered and followed by a one-on-one e-book reading by an experimenter. Experimenters included trained research assistants who were familiar with the books and trained on basic principles of dialogic reading. Experimenters read only the words on each page of the books, and did not provide any additional scaffolding. Each book reading session took around 20 minutes. Children were given the opportunity to choose which story they read each intervention day and were allowed to repeat books if they chose. Both stories followed the same basic structure. The condition of the e-book read to each child was randomly assigned. On days two and three, only the intervention was completed with the child. The fourth day consisted of the assessments only and did not include the intervention.

Measures

FAM Skill. Three test trial levels were included in the assessment testing FAM skills, and each trial item consisted of two sets of stars with 1-10 items in each set. The total surface area for each trial was controlled, such that the ratio of total object surface area on each side was the reverse of the ratio of the number of objects on each side in all incongruent test trials. For instance, on a trial with 3 stars on side A and 9 stars on side B (1:3 ratio), the total surface area of the stars on side A was three times greater than the total surface area of the stars on side B (3:1 ratio). Three simplified ratios were used for the number of stars in paired sets: 1:3, 2:3, and 1:2. Children were randomly assigned to begin with either the size trials or the number trials for the pre-switch level. In the size trials, the experimenter asked the child to choose the object set with the larger stars, while in the number trials, the experimenter asked the child to choose the object

set with the highest number of stars (See Figure 1 for sample trials). These tasks consisted of two practice trials and six test trials. Only the practice trials included feedback. For the post-switch level, children then were asked by the experimenter to attend to the dimension of magnitude that was not asked about in the pre-switch level. Again, feedback was given for the two practice trials, but not for the six test trials. At the pre- and post-switch levels, instruction was given on every trial (See Figure 1 for sample instruction, although exact wording varied slightly by trial). The final level was a mixed trial level in which the experimenter presented children with size trials and number trials in a randomized order and asked them to use the background color as a cue for which dimension to attend to. Children were reminded of the cue on some trials, but not on all trials (See Figure 1 for sample instruction). The mixed level consisted of two practice trials and twelve test trials. Several reminders of the cue were given throughout the game, but only the practice trials included feedback.

In the original FAM task (Fuhs et al., 2021), there was always one object set with more, yet smaller, items, while the other object set had fewer items, but they were larger. In other words, the object sets were always incongruent with regard to numerical and spatial magnitude. As a result, this task allowed for the possibility that children used a single-dimension strategy. To eliminate this possibility, the original FAM task was modified to include "check" trials throughout (Wagner et al., 2023). The "check" trials were congruent trials in which one of the object sets had both more numerous objects and larger objects. There were two "check" trials in the pre- and post-switch trials, and the mixed trials included four "check" trials. A child using a single-dimension strategy would be expected to achieve a high score on the test trials for the post-switch trial level, but answer the "check" trials incorrectly. Wagner et al. (2023) did not find evidence that a single-dimension strategy was used when averaging across the sample and found

very few children who appeared to use this strategy within trial levels. Additionally, FAM skill was found to still be predictive of math achievement, even when the subset of children who seemingly used a single-dimension strategy was removed from the analysis. In the current study, we used these check trials as well, but did not include them as part of the accuracy scores for test trial across levels. Consistent with prior studies, children scored significantly above chance on check trials within each trial level at both pre- and post-intervention (all *ps* < .05).

Number Line Estimation. Children's number line estimation skills were measured by the number-to-position task (Siegler & Opfer, 2003), which calls for awareness of both numerical and spatial magnitudes. Children were given a box with randomly selected numbers 1 through 10 and a number line that ran from 0 to 10. Children were instructed to place their finger on the number line where the boxed number would appear. The child answered correctly if they pointed exactly or near (1 number above or below) where the number would be on the number line. Scores were total correct responses.

Math Achievement. Overall math achievement was assessed as a standard score produced from the Number Sense subtest of the Woodcock Johnson IV Tests of Early Cognitive and Academic Development (WJ-IV ECAD) (Schrank & Dailey, 2014, 2015). This subtest is reliable and valid (Schrank et al., 2015) and requires students to use skills such as counting items on the page, estimating quantities of items in a picture, and identifying numbers to test their numerical skills. Standard scores were calculated and used in analyses.

Executive Functioning Skill. To evaluate students' EF skills, the Minnesota Executive Function Scale (MEFS) (Carlson & Zelazo, 2014) was used. This assessment is administered on an iPad and asks the child to sort cards into the correct box based on directions given to them by the examiner. This app was created based on the Dimensional Change Card Sort Task (DCCS;

Zelazo, 2006). The child either advances to the level above their previous level or goes to a level below, depending on the correctness of their choices. There are 7 levels that the child could attempt, with each level increasing in difficulty. Each level uses a different rule for how the cards should be sorted and can include tasks like sorting by size, color, and shape depending on the level of difficulty. As the levels increase, children must use EF skills to shift their focus between multiple characteristics in order to sort the cards correctly. A standard score was produced from the accuracy of the child's answers and from their response time, and this score was used in the analysis.

Books. This design utilized a pretest-intervention-posttest design with four one-on-one sessions with an experimenter at the child's school, and compared improvement in FAM skill across four conditions: size-to-number comparisons, mixed comparisons, traditional counting, and non-numerical (See Figure 2 for sample pages from each condition).

The size-to-number comparisons and mixed comparisons were experimental conditions, while the traditional counting and non-numerical conditions served as controls. The size-to-number condition books first included 10 question pages asking children to attend to the size of objects. Following the first set of question pages were 10 more question pages asking children to attend to the number of objects. Alternatively, the mixed condition included 20 question pages that alternated between asking children to attend to the size of objects and the number of objects in an arbitrary order. In both experimental conditions, each question page was followed by a feedback page either praising the child for answering correctly, or directing the child to the correct answer while emphasizing the incongruency between the size and number of objects. Including two experimental conditions enabled us to examine children's ability to switch from repeatedly focusing on one dimension of magnitude to repeatedly focusing on another (size-to-

number) and children's ability to continuously switch back and forth between dimensions (mixed). We expected that each of the experimental conditions would result in improvement on the FAM task at post-test; however, we did not have a prediction as to if one experimental condition would be more successful than the other.

In the traditional counting condition books, children were instructed to count the objects on each side of the page and choose which side had more objects. All objects in the traditional counting condition were the same size. This was repeated for 20 question pages, with feedback pages following each question page. In the non-numerical condition books, a single item appeared on the question page and children were instructed to name the object or animal that appeared on the pages and then describe either the color of the object/animal or the sound that the object/animal makes. Again, there were 20 question pages with feedback pages following each of them. The traditional counting condition control allowed us to rule out the possibility that it is simply the exposure to counting that drives the effect on math performance, and the non-numerical condition controlled for any influence that generally interacting with researchers may have had on performance.

Each child was randomly assigned to a condition within classroom so that each condition was approximately equally represented in each classroom, and the four one-on-one sessions consisted of experimenters reading to participants from different books of the same condition. Regardless of conditions, children had a choice between two tales that shared the same basic premise: Laura and Lamar visit a park, or Amaya and Oscar visit a zoo. In the first tale, the characters looked for a lost dog, while a penguin was used in the second. In each case, the story started with the animal going missing, and the characters had to complete tasks to find it. All conditions used the same objects and backgrounds and included the same number of pages. The

word count on each page was similar across conditions. In the traditional counting, size-to-number, and mixed conditions, the number of objects on each side of the page followed three different simplified ratios: 1:3, 2:3, 1:2. Each page followed these ratios using between 2 and 10 objects on each side of the page, to make sure that the numbers were within the range of numbers which preschool age children are familiar with. The size-to-number and mixed book conditions also varied the total area of objects on each side of each page based on these same three simplified ratios. The ratio of the total area on each side was the inverse of the ratio of total objects on each side. For instance, in a trial where the left side of the page contained 2 objects and the right side contained 4 objects (1:2 ratio), the objects on the left side had a total area that was twice the total area of the objects on the right side (See Figure 2 to view sample pages for each condition). The correct answer was distributed evenly across the right and left sides of pages.

Power Analysis and Exclusion Criteria. A power analysis conducted in G*Power for a 2 (timepoint: pre-intervention and post-intervention) x 4 (condition: counting, mixed, non-numerical, size-to-number) ANOVA, and a within-between interaction, revealed a necessary sample size of 84 participants for the study (21 participants per group). We assumed a medium effect size (η_{p^2} = .02), 2 timepoints, 4 between-participants groups, and a correlation between pre-intervention and post-intervention of r = .4 for the power analysis. We initially recruited 169 participants for the study, knowing that school absences during a short-term weeklong study would likely significantly reduce the final sample size. There were 116 participants in the final sample size (Size-to-Number condition n = 37, Mixed condition n = 25, Counting condition n = 30, Non-numerical condition = 24) who were present at all pre-intervention and post-intervention assessments and book reading sessions and had complete assessment data. We randomly

assigned children within each condition to either the number first FAM assessment or the size first FAM assessment (number first n = 61, size first n = 55). Participant exclusions were primarily due to school absences (n = 28), children not wanting to participate in one or more assessments (n = 13), developmental delays (n = 1), and technical errors with the online data collection system (n = 11).

Analytic Plan

Primary Analyses. Our primary research question was if children in the experimental FAM conditions significantly improved their FAM skills from pre-intervention to post-intervention. We did not have a prediction about whether one of the experimental FAM conditions would outperform the other experimental FAM condition. We conducted a repeated-measures ANOVA with timepoint (pre-intervention and post-intervention) as the within-participants factor and condition (size-to-number, mixed, counting, non-numerical) as the between-participants factor, controlling for age, gender, and EF skills. We analyzed planned contrasts for the condition by timepoint interaction that included a complex contrast testing for differences between the experimental (mixed and size-to-number) and control groups (counting and non-numerical) in their FAM task changes across timepoints, as well as simple contrasts comparing FAM changes across timepoints within each of the four conditions. We first examined the FAM mixed trials as our outcome of interest given that children were randomly assigned to complete different versions of the FAM task that varied in if the number or size trials were the pre-switch and post-switch trial levels. This mixed level was the trial level that was the same across all conditions. However, we also ran these analyses using the combined FAM post+mixed accuracy variable as this has been the variable used in prior correlational analyses, and we included FAM version as a covariate.

Exploratory Analyses. To explore potential differences in findings across number and size trials in the FAM task, we repeated our primary analyses, but with the number and size trials and pre- and post-intervention scores as the within-participants factor, controlling for FAM version and other covariates used in prior analyses. We also ran the same repeated-measures ANOVA with three other outcomes (global math achievement, number line estimation, and EF skills) to assess for potential transfer effects.

Results

Descriptive Statistics and Correlations

Descriptive statistics are presented in Table 1. Children scored below national norms on both EF skills and math achievement. FAM mixed trial level scores went up from pre-intervention to post-intervention; however, the pre- and post-switch FAM trial level scores went down slightly. We suspect that this occurred because these pre-switch and post-switch accuracy scores were collapsed across the two different FAM forms (size-number and number-size). Given that number and size trials are not necessarily equally difficult for all children, we did not expect FAM accuracy scores to increase on the overall pre- and post-switch levels (Wagner et al., 2023). In the next section, we explored performance across trial levels within each FAM version to further understand the interaction between trial level and FAM version. Correlations are presented in Table 2. As expected, pre-intervention and post-intervention scores for each measure were significantly correlated.

FAM Descriptive Analyses

We first examined if the FAM version children completed affected their performance on the FAM task pre-intervention. We conducted a 3 x 2 repeated measures ANOVA with FAM trial level (Pre-Switch, Post-Switch, and Mixed) and FAM version (number first or size first) as

predictors of FAM task accuracy. We used the Greenhouse-Geisser correction for sphericity. There was a significant interaction between trial level and version, $F(1.89, 215.49) = 5.49$, $p = .006$, $\eta_{p^2} = .05$. We explored this interaction by looking at performance across trial levels within each version. We found that scores for children who completed the number first FAM version did not significantly change from pre-switch (M = .95, SE = .02) to post-switch trials (M = .97, SE = .03, $t(60) = .94$, $p = .349$), but they did significantly decrease from both pre-switch trials to mixed trials (M = .64, SE = .03, $t(60) = 10.68$, $p < .001$) and post-switch trials to mixed trials ($t(60) = 11.49$, $p < .001$). For the size first FAM version, children's scores significantly decreased from pre-switch (M = .95, SE = .02) to post-switch trials (M = .83, SE = .03, $t(54) = 3.08$, $p = .003$), post-switch trials to mixed trials (M = .63, SE = .03, $t(54) = 5.01$, $p < .001$), and pre-switch trials to mixed trials ($t(54) = 8.73$, $p < .001$). We also examined potential differences across versions within each trial level. Within the pre-switch and mixed trial levels, children's scores did not differ by their FAM version condition (all *ps* > .05). Within the post-switch trial level, children scored significantly higher if they completed the number first version compared to the size first version ($t(114) = 3.58$, $p < .001$).

Primary Analyses

FAM Mixed Trial Level Outcomes. We conducted a 2 (timepoint: pre-intervention and post-intervention) x 4 (condition: counting, mixed, non-numerical, size-to-number) ANOVA predicting children's FAM mixed trial performance as this was the trial level that was identical across all study participants and was predicted to be the level most affected by condition. We also controlled for age, gender, and pretest EF skills. We predicted a significant interaction between timepoint and condition such that children in the experimental conditions (mixed, size-to-number) would significantly improve from pre- to post-intervention in their FAM skills

compared to control groups (counting, non-numerical). Overall, we did not find a significant interaction effect for timepoint x condition (Wilks' Lambda $F(3, 109) = .96$, $p = .41$, $\eta_{p^2} = .03$). We then tested our planned comparisons, as they were driven by our hypothesis that FAM ability can be improved through intervention. Prior studies examining the relevance of planned comparisons supported this decision by encouraging researchers to perform planned comparisons regardless of the outcome of the ANOVA analysis, especially when those comparisons are supported by previous findings or theory (Ruxton & Beauchamp, 2008)

First, we examined pre-intervention to post-intervention change in FAM task performance within each condition, and then compared the two experimental conditions to the two control conditions. Children significantly improved their FAM task performance from pre-intervention to post-intervention within the two experimental groups, mixed (Wilks' Lambda $F(1, 109) = 8.65$, $p = .004$, $\eta_{p^2} = .07$) and size-to-number (Wilks' Lambda $F(1, 109) = 4.77$, $p = .031$, $\eta_{p^2} = .04$) (See Figure 3). Children did not significantly improve their FAM task performance in the two control group conditions, counting (Wilks' Lambda $F(1, 109) = .54$, $p = .463$, $\eta_{p^2} = .01$) and non-numerical (Wilks' Lambda $F(1, 109) = 1.77$, $p = .186$, $\eta_{p^2} = .02$) (See Figure 3). These findings also held when comparing the two experimental conditions to the two control conditions such that children in the two experimental conditions significantly improved their performance (Wilks' Lambda $F(1, 111) = 12.90$, $p < .001$, $\eta_{p^2} = .10$) (See Figure 4) while children in the control conditions did not (See Figure 2) (Wilks' Lambda $F(1, 111) = 2.11$, $p = .150$, $\eta_{p^2} = .02$) (See Figure 4).

FAM Post + Mixed Trial Level Outcomes. We also conducted the same analyses with the exception of the outcome variable being the FAM Post+Mixed task performance as this variable has been used in prior FAM studies (Fuhs et al., 2021). We included FAM version as a covariate

given that children completed either the number or size trials at post-switch. Overall, we did not find a significant interaction effect for timepoint x condition (Wilks' Lambda $F(3, 105) = .98$, $p = .591$, $\eta_p^2 = .02$). We then examined our planned contrasts. First, we examined pre-intervention to post-intervention change in FAM task performance within each condition, and then compared the two experimental conditions to the two control conditions. All of these contrasts yielded null findings, suggesting that the intervention changes we observed were due primarily to differences within the FAM mixed trial level rather than the FAM post-switch trial level.

Potential Transfer Effects. We examined pre-intervention to post-intervention changes in overall math achievement, number line estimation, and EF skills. We continued to covary age and gender, and for the math achievement and NLE analyses, we covaried pretest EF skills as a general skill variable. For overall math achievement, we did not find a significant interaction between timepoint and condition (Wilks' Lambda $F(3, 109) = .61$, $p = .605$, $\eta_p^2 = .02$), nor did we find significant differences within either experimental or control groups in their pre- to post-intervention FAM score changes. For NLE scores as the outcome variable, we did not find an overall interaction effect between timepoint and condition (Wilks' Lambda $F(3, 109) = 1.82$, $p = .618$, $\eta_p^2 = .02$). For planned contrasts, we did not find any significant differences across specific conditions in change from pre-intervention to post-intervention. Combining the two experimental conditions and the two control conditions for the complex contrast, we found a significant difference between the changes in NLE scores of experimental and control groups (Wilks' Lambda $F(1, 111) = 5.17$, $p = .025$, $\eta_p^2 = .04$) such that the experimental conditions overall significantly improved their NLE scores from pre- to post-intervention (NLE Pre-Int total = 1.60, SE = .26; NLE T2 total = 2.07, SE = .29; Wilks' Lambda $F(1, 111) = 4.40$, $p = .038$, $\eta_p^2 = .04$), while the control conditions did not (NLE Pre-Int total = 2.33, SE = .28; NLE Post-Int total

= 2.07, SE = .21; Wilks' Lambda $F(1, 111) = 1.37$, $p = .245$, $\eta_{p^2} = .01$) (See Figure 5). Finally, we examined condition differences on EF skills changes from pre- to post-intervention. We found a non-significant interaction effect for timepoint x condition (Wilks' Lambda $F(3, 110) = 2.42$, $p = .070$, $\eta_{p^2} = .06$). Planned contrasts revealed that only the non-numerical control condition improved their EF skills from pre- to post-intervention (EF Pre-Int = 92.70, SE = 1.71; EF Post-Int = 96.38, SE = 1.95; Wilks' Lambda $F(1, 110) = 6.32$, $p = .013$, $\eta_{p^2} = .05$).