JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
2016, VOL. 9, NO. 4, 577–606
http://dx.doi.org/10.1080/19345747.2016.1191575

INTERVENTION STUDIES

Effects of Tutorial Interventions in Mathematics and Attention for Low-Performing Preschool Children

Marcia A. Barnes[a], Alice Klein[b], Paul Swank[c], Prentice Starkey[b], Bruce McCandliss[d], Kylie Flynn[b], Tricia Zucker[c], Chun-Wei Huang[e], Anna-Mária Fall[a], and Greg Roberts[a]

ABSTRACT

Two intervention approaches designed to address the multifaceted academic and cognitive difficulties of low-income children who enter pre-K with very low math knowledge were tested in a randomized experiment. Blocking on classroom, children who met screening criteria were assigned to a Math + Attention condition in which the Pre-Kindergarten Mathematics Tutorial (PKMT) intervention was implemented (4 days/week for 24 weeks) in addition to 16 adaptive attention training sessions, a Math-Only condition using the PKMT intervention, or a business-as-usual condition. Five hundred eighteen children were assessed at pretest and posttest. There was a significant effect of the PKMT intervention on a broad measure of informal mathematical knowledge and a small but significant effect on a measure of numerical knowledge. Attention training was associated with small effects on attention, but did not provide additional benefit for mathematics. A main effect of state on math outcomes was associated with a stronger, numeracy-focused Tier 1 mathematics curriculum in one state. Findings are discussed with respect to increasing intensity of math-specific and domain-general interventions for young children at risk for mathematical learning difficulties.

KEYWORDS
randomized controlled trial
pre-kindergarten
mathematics
attention
math difficulties

Mathematical knowledge at school entry is the strongest predictor of later academic achievement (Claessens, Duncan, & Engel, 2009; Duncan et al., 2007). Children with low number skills at kindergarten entry are at high risk for low math achievement over the early elementary school grades (Jordan, Kaplan, Ramineni, & Locuniak, 2009). This risk is greatest for low-income preschool and kindergarten children who possess less extensive mathematical knowledge than their middle-class peers in preschool and kindergarten (Ginsburg & Russell, 1981; Griffin, Case, & Siegler, 1994; Jordan, Huttenlocher, & Levine, 1992; Saxe, Guberman, & Gearhart, 1987; Starkey & Klein, 1992, 2008).

Longitudinal studies show that children who start low in mathematics continue to struggle in the early elementary years and into high school (Bodovski & Farkas, 2007; Hanich

CONTACT Marcia A. Barnes marcia.barnes@austin.utexas.edu Department of Special Education, University of Texas at Austin, Austin, TX 77030, USA.
[a]University of Texas, Austin, Texas, USA
[b]WestEd, Oakland, California, USA
[c]University of Texas Health Science Center at Houston, Houston, Texas, USA
[d]Stanford University, Stanford, California, USA
[e]WestEd, San Francisco, California, USA

& Jordan, 2001; Jordan, Kaplan, Oláh & Locuniak, 2006; Shalev, Manor, & Gross-Tsur, 2005). Specifically, the math achievement of low-income children between first and third grades lags behind that of their middle-income peers and these lags are mediated by incoming kindergarten number competence (Jordan et al., 2009), suggesting the need for interventions for low-income preschoolers that address SES-related gaps in mathematical knowledge.

The reason for the early SES-related gap in mathematical knowledge stems, at least in part, from differential levels of support for early mathematical development that children receive in their early learning environments, both at home (Hart & Risley, 1995; Saxe et al., l987; Starkey & Klein, 2008) and at school. Children from low-income families, for example, receive less support for mathematical development in their preschool classrooms. Mathematics instruction often is not systematic, and preschool teachers have traditionally received little training in early childhood mathematics and they are not familiar with pre-K mathematics standards (Copley, 2004; Copley & Padron, 1999; Farran, Silveri, & Culp, 1991).

In order to address the SES-related gap in early mathematical knowledge, several researchers have developed curricula or interventions to support young children's mathematical learning in public preschool classrooms (Casey, Kersh, & Young, 2004; Clements, Sarama, Spitler, Lange, & Wolfe, 2011; Griffin, 2004; Lewis Presser, Clements, Ginsburg, & Ertle, 2015; Ramani & Siegler, 2008; Sophian, 2004; Starkey, Klein, & Wakeley, 2004). These preschool mathematics curricula are essentially Tier 1 classroom interventions developed for use with the general population of economically disadvantaged preschool children to prepare them for school mathematics. Overall, these curricular interventions have been shown to improve young children's mathematical knowledge.

Pre-K Mathematics (Klein & Starkey, 2004) is a Tier 1 intervention that has been rigorously evaluated and found to significantly increase low-income children's mathematical knowledge. This early math intervention uses small-group activities with concrete materials to help children build a broad foundation of informal mathematical knowledge and thereby prepare them for more abstract (formal) mathematical thinking in elementary school. Pre-K Mathematics has been shown to be effective in multiple randomized controlled trials, including a study where it was implemented at scale (Klein, Starkey, Clements, Sarama, & Iyer, 2008; Starkey & Klein, 2012; Starkey, Klein, DeFlorio, & Swank, 2016), with effect sizes ranging from .58 to .83 across these studies. Despite the effectiveness of this intervention for low-income preschool children in general, there appears to be a subgroup of children who receive the math curriculum but continue to show low growth in mathematics over the school year. Specifically, it was found that children who score in the lowest quartile on measures of mathematical knowledge at the beginning of pre-K scored lower at the end of pre-K than the other intervention children (Starkey & Klein, 2012). In fact, this subgroup of children scored lower than some of the control children who did not receive Pre-K Mathematics.

Early Tier 1 math interventions are intended to improve low-income preschool children's mathematical learning and to prepare them for mathematics instruction in kindergarten. However, it is clear that there is a subgroup of pre-K children who demonstrate low response to generally effective Tier 1 math interventions. These children exit pre-K well below kindergarten readiness benchmarks for mathematics (NAEYC/NCTM, 2002), and are therefore at risk for significant math difficulties throughout elementary school. Children entering and exiting kindergarten below the 10th percentile in math, for example, have a 70% chance of scoring below the 10th percentile five years later (Morgan, Farkas, & Wu, 2009).

The need to find solutions for children with low response to high-quality interventions is a topic of considerable urgency and interest to the special education research and practice community. Hypotheses about the sources of low responsiveness and solutions to increase learning in these children include: significantly intensifying interventions; increasing the comprehensiveness of taught skills and strategies; explicitly teaching for transfer; and addressing the cognitive and linguistic limitations of students with learning disabilities (e.g., Fuchs & Fuchs, 2014; Fuchs, Fuchs, & Vaughn, 2014). In keeping with these ideas, the current study attempted to: (a) significantly increase the intensity of a mathematics intervention for preschool children at high risk for math difficulties and (b) address cognitive weaknesses by combining an intensive math intervention with an attention intervention. The rationale for these approaches to address the needs of pre-kindergarten children at high risk for math difficulties is reviewed below.

Systematic reviews of the literature on math interventions for students with math learning difficulties highlight the importance of providing instruction that incorporates particular design features: increases the intensity of instruction in mathematics in addition to Tier 1 instruction; provides explicit, systematic instruction that integrates developmental research in mathematics with principles of direct instruction; provides opportunities for cumulative review; teaches mathematical concepts to mastery ensuring that foundational skills needed to understand higher level concepts are mastered before moving on to the next level; uses concrete and visual representations including manipulatives; provides tutor scaffolding for learning as well as emotional support; and tracks children's understanding of each mathematical concept, adjusting the difficulty of concepts in relation to child knowledge (Fletcher, Lyon, Fuchs, & Barnes, 2007; Gersten et al., 2009).

Furthermore, higher tiers of instruction typically employ small group pull-out interventions that supplement rather than replace Tier 1 instruction (e.g., Bryant, Bryant, Gersten, Scammacca, & Chavez, 2008; Powell et al., 2015; review in Fletcher et al., 2007; Vaughn, Denton, & Fletcher, 2010). To our knowledge, these more intensive approaches to math instruction for at-risk children have not been tested at the pre-K level. In the current study, a small-group tutorial-based intervention for children entering pre-K with very low math knowledge was created using the design features of effective interventions for children with math difficulties reviewed above.

In addition to testing the effect of a tutorial-based intervention for low-performing pre-K children, the current study also investigated whether an intervention that addressed cognitive weaknesses in attention, when combined with an intensive math intervention, would have any synergistic effects on math outcomes. This approach is based on what is known about the relation of math and attention in children with learning disabilities, research on the relation of attention and mathematics from neurobiological and behavioral studies, and the central importance of attention for young children's learning.

Attention abilities in the preschool years and at the beginning of kindergarten are a significant predictor of later academic achievement in reading and mathematics (Duncan et al., 2007; McClelland, Acock, Piccinin, Rhea, & Stallings, 2013). Strong phenotypic and genetic correlations between attention and math have been found across the distribution of mathematical achievement, and these associations are not fully accounted for by other academic or cognitive variables (Greven, Kovas, Willcutt, Petrill, & Plomin, 2014). For children with math disabilities, the association of attention difficulties and mathematics achievement is significant and large even among those children who do not meet diagnostic criteria for

ADHD (reviewed in Tannock, 2013). Furthermore, the severity of the math learning disability is related to the presence of inattention, and inattention predicts difficulties in both conceptual and procedural aspects of math (Cirino, Fletcher, Ewing-Cobbs, Barnes, & Fuchs, 2007; Raghubar et al., 2009; Tannock, 2013). Difficulties in attention have also been associated with less response to intervention (Miller et al., 2014; Rabiner & Malone, 2004). These findings suggest that attention is not simply a cognitive correlate of mathematics, but is importantly related to math performance in children with and without difficulties in mathematics.

Attention is referred to as a "hub" cognitive domain in that it performs a gatekeeping function for the acquisition of skills across many areas, particularly for young children (Garon, Bryson, & Smith, 2008; Posner & Rothbart, 2007; Wass, Scerif, & Johnson, 2012). For example, aspects of self-regulation associated with attention in pre-K children predict later math performance over the early school years even after controlling for several demographic, cognitive, and academic variables (Blair, Ursache, Greenberg, & Vernon-Feagans, 2015). Finally, attention training may be particularly effective for young children (review in Wass et al., 2012). Studies with 4- and 5-year-old children show that such training has both immediate and delayed effects on the neural networks for attention (Rueda, Checa, & Cómbita, 2012).

Although there has been interest in the potential for cognitive interventions to improve academic outcomes in high-risk populations, there is currently little intervention research that explicitly focuses on the multifaceted nature of learning difficulties and that attempts to provide interventions for both academic and cognitive skills. Most cognitive intervention research has been conducted with typically developing children and does not investigate the combined effects of cognitive plus academic skills interventions; the type of cognitive training varies quite significantly between studies; the age of the children in the interventions varies; interventions are relatively brief; and academic outcomes are not always or often the outcomes of interest. Recent meta-analyses on the effect of cognitive interventions such as working memory and other executive function training on academic skills have not yielded particularly positive findings (Jacob & Parkinson, 2015; Melby-Lervåg & Hulme, 2013). However, there are very few high-quality experimental studies of the effects of combined cognitive and academic interventions on mathematics with randomized designs and fewer still that are specifically relevant for young children at high risk for learning difficulties (Jacob & Parkinson, 2015; Wass, 2015; but see Kroesbergen, van't Noordende, & Kolkman, 2014).

In contrast to the mostly negative findings regarding effects of working memory training on academic achievement for older typically developing children, findings for attention training have been more positive (Peng & Miller, 2016). Furthermore, effects of attention training are moderated by type of population studied and age, such that effects are larger for children with attention difficulties and for younger children (Peng & Miller, 2016; Wass et al., 2012). These considerations led us to question whether the pre-kindergarten year might be an optimal developmental and instructional time for testing the combined effects of domain-specific mathematics and domain-general attention interventions for preschool children with very low math knowledge.

Our research questions and hypotheses were as follows:

1. Does an intensive early math intervention significantly improve the math knowledge of low-performing pre-K children as compared to business-as-usual in pre-K

classrooms? We predicted that the groups receiving the PKMT intervention would outperform the business-as-usual group on assessments of math knowledge.

2. Does attention training improve the attention abilities of these pre-K children compared to children who receive only the math intervention or business-as-usual? We predicted that the group receiving attention training would outperform the business-as-usual group and the group receiving only the PKMT intervention on attention outcomes.
3. Does attention training combined with an intensive early math intervention improve the math knowledge of these children compared to children who receive only the math intervention? To the extent that attention facilitates mathematical learning through improved ability to focus attention on task-relevant information and ignore task-irrelevant information during instruction and problem-solving practice, we predicted that the group receiving the combined intervention would have higher scores on assessments of math knowledge than the group receiving only the PKMT intervention.

Method

Participants

States and Preschool Programs

This study was conducted in state pre-K programs in two states, Texas and California, in order to recruit a sufficiently large sample of very low-performing pre-K children for this randomized controlled trial. In Texas, the study was implemented in 16 public schools within an urban school district, Houston Independent School District (HISD). The district's core mathematics curriculum was Scholastic Frog Street. This Tier 1 math curriculum was supplemented with additional materials and professional development delivered by the early childhood program staff at the district-level during a daily "math block" that included whole-group instruction followed by mathematics' workstations and/or small-group activities. In California, the study was implemented in three school districts or local educational agencies in the greater Bay Area. Specifically, 10 public schools within the San Jose Unified School District (SJUSD), 14 schools within the West Contra Costa Unified School District (WCCUSD), and five state pre-K centers within the Santa Clara County Office of Education (SCCOE) participated in the study. Growing with Math was the core (Tier 1) mathematics curriculum adopted by SJUSD and SCCOE, while Everyday Math was the core curriculum reported by WCCUSD.

Classrooms

A total of 95 pre-K classrooms participated in the study across both states in the 2012–13 (Cohort 1) and 2013–14 (Cohort 2) school years. In Texas, there were 49 classrooms and teachers in the study with all classrooms having full-day programs. Thirty-six teachers reported that they provided English-only instruction and 13 teachers reported providing most instruction in Spanish. These mostly Spanish instruction classrooms were classified as part of the district's transitional bilingual model (90% Spanish/10% English in pre-kindergarten). In California, there were 46 classrooms and teachers in the study. The majority of classrooms in California followed a half-day schedule, with 36 part-day classrooms and

10 full-day classrooms included in the sample. English was the primary language of instruction in all the California classrooms.

Children

Children were selected for participation in this study through a multistep process that included determining their age and language eligibility, screening for very low math performance, and obtaining parental consent. The first step was to obtain date of birth information for all children in participating classrooms to determine their age-eligibility to attend traditional kindergarten the following school year. In order to be included in the study, children had to be 4 years of age before the first intervention session. All children who were age-eligible were then assessed on the Math Screening measure (see Measures). Across both cohorts and states, 1,700 children were screened. In addition, children who were monolingual in a language other than English or Spanish were excluded, because they would not be able to participate in the assessments. Finally, children who qualified for the study based on their Math Screening performance were included if they also met the inclusion and exclusion criteria detailed in the following section.

Inclusion and Exclusion Criteria. Children with autism spectrum disorders and significant behavioral difficulties were excluded, as were children with severe cognitive delay. We did not exclude children based on the presence of any other neurodevelopmental or learning disorders. Because we had only a nonverbal measure of cognitive ability and no measure of verbal intelligence or adaptive function needed to identify intellectual disability, we retained children in our sample who scored below a standard score of 70 on the K-BIT Matrices subtest. Overall, 10% of the sample scored below 70. The average K-BIT Matrices score for the sample was 91.3 (Table 1). Table 1 summarizes the demographic characteristics of the study sample. Race/ethnicity information was obtained in Texas from school records or parents, and in California from parents.

Consent forms were distributed to parents of all eligible children based on their math screening scores and the inclusion and exclusion criteria above. Because the design of the study required equal numbers of children in each of the three conditions from each classroom resulting in three, six, or nine participants per classroom, consents were accepted until the requisite number of children per classroom was obtained. Occasionally, more students within a classroom were consented than we were able to include. In these cases, we used a randomized process to select study participants. Most classrooms provided six students, with fewer classrooms providing three or nine students.

All children who participated in the study were from low-income families and were income-eligible to attend state pre-K programs in their respective states. A total of 541 children were included in the study at pretest across cohorts, 277 in Texas and 264 in California. Five-hundred eighteen children were assessed at posttest, 261 in Texas, and 257 in California. The reasons for not being able to assess children at posttest were as follows in order of occurrence: withdrawal from school; relocation out of region or state; withdrawal of consent due to change in guardianship.

Attrition. Overall attrition across both states and cohorts was 4%. Attrition was 6% in Texas, and 3% in California. Attrition did not differ across conditions.

Table 1. Demographic characteristics of pretest sample.

		Overall	$M + ATT$	Math only	BAU
Sample *N*					
	Total	541	181	180	180
	TX	277	93	92	92
	CA	264	88	88	88
Gender (% female)					
	Total	46.7	44.8	50.0	47.8
	TX	45.5	45.2	43.5	48.9
	CA	48.5	44.3	56.8	46.6
Race/ethnicity (%)					
Hispanic	Total	71.7	75.7	70.6	68.9
	TX	54.2	57.0	53.3	52.2
	CA	90.2	95.5	88.6	86.4
African American	Total	17.9	17.1	17.2	19.4
	TX	31.8	31.2	30.4	33.7
	CA	3.4	2.3	3.4	4.5
Asian American	Total	1.7	0.6	2.2	2.2
	TX	0.7	1.2	0.0	1.2
	CA	2.7	0.0	4.5	3.4
Caucasian	Total	2.2	1.7	3.3	1.7
	TX	3.2	2.2	4.3	3.3
	CA	1.1	1.0	2.3	0
Mixed/other	Total	3.7	2.2	3.3	5.0
	TX	4.7	3.2	5.4	5.4
	CA	2.7	1.1	1.1	5.7
Unknown	Total	2.8	2.8	3.3	2.2
	TX	5.4	5.4	6.5	4.3
	CA	n/a	n/a	n/a	n/a
Child age in years, *M* (*SD*)					
	Total	4.50 (0.27)	4.50 (0.27)	4.53 (0.27)	4.46 (0.26)
	TX	4.57 (0.28)	4.56 (0.28)	4.63 (0.27)	4.53 (0.28)
	CA	4.42 (0.23)	4.43 (0.23)	4.42 (0.23)	4.40 (0.23)
KBIT pretest, *M* (*SD*)					
	Total	91.25 (14.29)	92.14 (13.96)	92.00 (13.60)	89.61 (15.22)
	TX	90.78 (14.46)	92.10 (14.01)	92.19 (13.34)	88.04 (15.82)
	CA	91.51 (14.22)	92.16 (14.01)	91.90 (13.82)	90.47 (14.91)

Note. $M + ATT$ = math + attention condition; Math only = math only condition; BAU = business-as-usual condition; CA = California; TX = Texas; KBIT = KBIT Matrices subtest administered only at pretest.

Research Design

The intervention effects of interest are increases in mathematical knowledge for pre-K children who are very low performing in math at the beginning of the school year and receive an intensive, tutorial math intervention by itself or in combination with an attention-training program. Because the smallest educational unit for which a tutorial math intervention could be implemented is the student within the classroom, the student was chosen as the appropriate unit of randomization for this experiment. Blocking on classrooms, equal numbers of children within each classroom who met the screening criteria (see below) and received parent consent were randomly assigned to one of three conditions: (a) Math + Attention or $M + ATT$ condition, in which children received both the tutorial math intervention and attention training; (b) Math only or *M only* condition, in which children received the tutorial math intervention and a control activity ("books on tape" described below) for the same amount of time as the attention program provided to children in Condition 1; or (c) A business-as-usual condition, in which children did not receive either the tutorial math intervention or attention training. Note that children in all three conditions continued to receive their regular Tier 1 math instruction from their classroom teacher and that a trained tutor delivered the math and attention interventions outside of the classroom. Thus, by randomly assigning children to conditions within classrooms, the design equated the effect of Tier 1 classroom math instruction across conditions.

Interventions

PKMT Intervention

The tutorial math intervention used in this study, Pre-K Mathematics Tutorial (PKMT), was designed to provide intensive instruction on a core set of math concepts and skills to help low-performing children build a strong foundation of informal mathematical knowledge. The intervention is adapted from an effective Tier 1 mathematics curriculum, Pre-K Mathematics (Klein & Starkey, 2004). The PKMT program comprises 20 math activities that use concrete materials to engage pre-K children and to support their mathematical learning of concepts related to number, arithmetic, space, geometry, and measurement. The mathematical content of these activities is based on cognitive developmental research (see Bisanz, Sherman, Rasmussen, & Ho, 2005; Ginsburg, Klein, & Starkey, 1998 for reviews) and, in keeping with the recommendations of the National Council of Teachers of Mathematics Focal Points (2006), more than half of the PKMT math activities focus on Number and Operations.

The structure of each math activity is designed to be sensitive to the learning needs of low-performing pre-K children. The activities are scripted to enable tutors to provide explicit (e.g., modeling) and systematic instruction, and to scaffold children when they encounter difficulty. A downward (less challenging) extension is provided for children who are not ready for the math activity, and an upward extension is included for children who succeed easily on the activity. PKMT is unique in its focus on providing instruction on a core set of concepts and skills that would be expected at the beginning of pre-K, but are not yet developed in this very low-performing population. As such, the first half of the PKMT intervention involves foundational math activities (e.g., learning the sequence of number names) and the second half progresses to more advanced pre-K math activities (e.g., matching concrete sets to number names and numerals). Finally, PKMT incorporates a number of the key characteristics that have been found to be effective in interventions for children at risk for mathematical difficulties: (a) increased intensity of math instruction beyond the Tier 1 curriculum; (b) explicit, systematic instruction that integrates developmental research about mathematics with principles of direct instruction; (c) cumulative review; (d) teaching to mastery; (e) scaffolding for learning and for providing emotional support; and (f) progress monitoring to track children's understanding of each math concept and adjust instruction to children's knowledge.

The PKMT activities were delivered by a trained tutor to pairs of children in their dominant language outside of the classroom. The intervention was implemented over 24 weeks, with a new math activity introduced each week (20 weeks) and four review weeks scheduled over the course of the intervention. Children participated in tutorial math sessions four days a week (Monday–Thursday) for approximately 15–20 minutes each day, and Fridays were reserved for review and/or makeup sessions for children who were absent during the week. The tutors worked with pairs of children from the same classroom, one from the Math-Only condition and one from the Math + ATT condition, to ensure that the PKMT intervention was implemented with equivalent fidelity across the two treatment conditions. Moreover, they implemented the PKMT intervention according to a carefully sequenced curriculum plan that introduced foundational math activities before progressing to more challenging pre-K math activities. Finally, in order to monitor children's progress over the course of the intervention, tutors used Assessment Record Sheets to record the child's understanding after each dosage (presentation) of an activity and a Math Mastery document to track the child's mastery across the set of PKMT math activities.

Attention Intervention (ATT)

In addition to the PKMT intervention, the students in the Math + ATT group participated in a computerized, attention-training intervention consisting of two types of attention-training games—Vigilance and Conflict (Rueda et al., 2012; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005). This particular game-based attention intervention was chosen because it is associated with improvements on measures of attention for 4- to 5-year-old children along with associated changes in the neural systems that subserve attention (Rueda et al., 2012). The Vigilance games require children to sustain attention and respond to a stimulus by pressing the spacebar on a laptop computer to feed different animals (a frog, an anteater, and a bear) their favorite food (flies, ants, fish). For example, Francis the frog loves to eat flies that are trapped inside a jar. When one flies out, the child must "catch" the fly as quickly as they can after the fly exits the jar. The game adapts to the child's speed in "catching flies" between the time they leave the jar and the child presses the button. Feedback is provided on each trial in the child's first language, including corrective feedback if the child does not wait for the fly to exit the jar before pressing the button (trains inhibition of a pre-potent response) and also if the child waits too long after the fly exits the jar before pressing the button (trains child to be vigilant to quickly detect targets). Over the training sessions, "catch times" (the length of time it takes for the child to press the button when the fly leaves the jar) become faster as the program adapts to child responses. The Conflict game requires children to make a choice between two stimuli based on rules that change across trials in the game. In this game, which also adapts to child progress, two monsters—KiKi and Booba—love to eat different foods. During the day (signaled by a sun in the sky), KiKi eats red flowers and trucks and Booba eats blue flowers and trucks; but at night (signaled by a moon in the sky), KiKi eats all the flowers and Booba eats all the trucks, regardless of their color. The game begins with the daylight phase and changes in difficulty and complexity when the nighttime phase is introduced. The computer program switches between daytime and nighttime and increases in accuracy and response times are used to detect when the child should move through higher levels of the games. The game is considered to tap executive attention because it requires the ability to regulate one's attention by focusing on task-relevant information, ignoring task-irrelevant information and inhibiting pre-potent responses, and switching flexibly between response rules.

The games were presented on Dell laptops with 13-inch displays. Children were given 16 sessions of ATT, with eight sessions each of computerized Vigilance and Conflict games over the course of 16 weeks. Half of the participants received the Vigilance sessions first and half received the Conflict sessions first. Children played the ATT games for 8 minutes per session one time per week (typically Fridays when PKMT was not implemented). These were one-on-one sessions and during the session tutors could redirect the child to attend to the task if children were engaging in off-task behavior that disrupted their playing of the games. At the end of each 8-minute session, the tutor awarded the appropriate number of "attention reward" stamps to the child based on his or her time spent on-task to improve motivation and engagement during ATT. Based on pilot data, the ATT sessions were relatively short and occurred on the week day when the PKMT intervention was not being implemented. Issues regarding the low intensity of the attention intervention are further addressed in the "Discussion" section.

Books on Tape Control Activity

To control for effects of additional tutoring time and attention from the tutor (i.e., Hawthorne effects), the students assigned to the *M only* condition met with their tutor for a books on tape session one time per week for 16 weeks. During these sessions, students listened to 8 minutes of a prerecorded text while looking at the book. To improve motivation, students in this group also received stamps in their stamp books based on the time spent on task. Students listened to books on tape in their preferred language (English or Spanish).

Tutor Training

Tutors were hired based on having experience working with young children either in a teaching or assessment capacity. Most of the tutors had acquired a bachelor's degree. Tutors without advanced postsecondary degrees were required to have teaching experiences at the prekindergarten level. Bilingual tutors were hired based on these qualifications, but were also required to pass certification on the PKMT and ATT interventions in both English and Spanish. The tutors underwent a rigorous training and certification process in order to be prepared to deliver the PKMT and attention interventions with fidelity. They received four days of training on the PKMT intervention, two days in the early fall on the foundational activities (i.e., the first half of the curriculum) and two days in the winter on the pre-K activities (i.e., the second half of the curriculum). Another full day of training was provided on the ATT intervention in the fall. After these trainings, tutors were given additional opportunities for guided practice with the professional development (PD) specialist on both the PKMT activities and the ATT games. Then, all tutors were required to pass a two-step certification process: (a) staff certification in the lab to be sure that they could execute both interventions with fidelity and (b) field certification with a child at a preschool under the observation of the PD specialist research assistant to ensure accurate delivery of the PKMT intervention (i.e., implementing with fidelity, providing appropriate scaffolding, accurately recording progress monitoring) and ATT programs (i.e., correctly training the children to use the computers and introducing the first session for each game, running subsequent sessions, correct usage of the reward stamps). Ongoing support was provided during the course of the intervention through weekly tutor calls where tutors discussed their children's progress and challenges on the math activities, and additional strategies for scaffolding and review were provided. Tutors also reported the progress that children made on the attention-training programs from the previous week (e.g., catch time scores on the Vigilance games and student on-task reward stamps) and troubleshooting was provided as needed.

Fidelity

Fidelity of implementation for both the PKMT and ATT interventions are reported below. The professional development specialist in each state conducted six fidelity observations of each tutor who implemented the PKMT intervention with children during the pre-K year. The following dimensions of fidelity were observed and rated: following a schedule, setup of materials, delivery of the basic activity, developmental adjustments to the basic activity, and assessment of children's mastery. The overall mean fidelity scores for the PKMT intervention were very high, .98 in Texas and .99 in California. Similarly, trained project staff in each state conducted approximately four fidelity observations of each tutor to evaluate the quality of the ATT intervention. Tutors were scored on how accurately they implemented the intervention as intended on a 3-point scale, including their ability to follow scripts, give corrective prompts as needed, provide the correct number of stamps, and keep accurate records. The overall mean fidelity score for the ATT activities was also very high at .98. In addition, based on the computerized data for the ATT sessions, the number of minutes logged for each child for each game was as intended (128 minutes across 16 sessions).

Tier 1 Math Instruction in Classrooms

The Early Mathematics Classroom Observation (EMCO) was used to collect quantitative and qualitative data on Tier 1 mathematics instruction in participating classrooms. Classroom observers were trained and certified on the EMCO prior to the onset of classroom observations. Observations were conducted in the fall of the school year. Inter-rater reliability on the total minutes of math (MOM) observed in the classroom was .98.

EMCO observations were conducted to determine whether the intended Tier 1 mathematics curriculum was being implemented in participating classrooms. These observations also identified types of practices teachers use to teach math and the amount of time these practices were used. Observations were made of all teacher–participant activities in which there was mathematical content. Math activities were categorized as *focal*, in which mathematics learning was a primary goal of the activity, or as *embedded*, in which mathematics learning was an incidental goal embedded in a nonmathematical activity. The settings of these activities were categorized as whole-group settings (the entire class and teacher were engaged in the activity simultaneously) or as small-group settings (a subset of children and a teacher were engaged in the activity). The duration and number of children present during each teacher–participant math activity were recorded. Data were then scored for the total MOM instruction that each child received, on average, during an observation session as well as the type of math instruction: (a) focal, whole-group MOM, (b) embedded, whole-group MOM, (c) focal, small-group MOM, (d) embedded, small-group MOM.

The EMCO observations confirmed that participating teachers were implementing Tier 1 mathematics curricula in their classrooms. However, implementation of these mathematics curricula varied considerably by state and by type of math activity (Table 2). Significantly more time was devoted to mathematics instruction by Texas teachers than by California teachers, in regard to both total MOM, $F(1, 92) = 36.84$, $p < .0001$, and focal whole-group MOM, $F(1, 92) = 57.60$. $p < .0001$. We consider the implications of these state differences in Tier 1 math instruction in the "Discussion" section.

Measures

The math screening measure, a nonverbal measure of cognitive ability (KBIT-2), and the principal math and attention measures used in the impact analyses are described below. Measures of other cognitive abilities examined in this study (working memory, phonological awareness, acuity of approximate number system) are not described or reported in this paper.

Table 2. EMCO mean minutes of math (MOM) provided by teachers in pre-K classrooms by site.

State	Focal small group MOM	Focal whole group MOM*	Embedded small group MOM	Embedded whole group MOM	Mean total MOM (*SD*)*
CA	3.3	6.5	0.6	2.0	12.4 (11.1)
TX	2.3	22.3	0.0	2.7	27.3 (12.7)

Note. $^{*}p < .0001$.

Math Screening Measure

All age-eligible pre-K children in participating classrooms were screened at the beginning of the school year on a measure comprising three tasks from the Child Math Assessment (Starkey & Klein,, 2012). The math screening measure was designed to identify a group of low-performing children in the pre-K year who were at risk for low math achievement in kindergarten. The specific tasks were selected through a series of logistic regression analyses using two data sets from a multistate, randomized controlled trial of the Pre-K Mathematics intervention (Starkey & Klein, 2012). The first data set included 744 pre-K children who received the full Child Math Assessment (CMA) at pretest, and the second data set included a longitudinal follow-up assessment of 618 children on the TEMA-3 (Ginsburg & Baroody, 2003) at the end of kindergarten. The logistic regression analyses were conducted to: (a) identify which CMA tasks in pre-K best predict kindergarten performance on the TEMA-3 below the 25th percentile (indicator of low math achievement), and (b) determine the predicted probability cutoff on the screener measure that would yield a correct classification or hit rate of 70%. In the logistic regression analyses, the total correct score on each CMA task served as the predictor, and the percentile rank (PR) performance on the TEMA served as the outcome variable with a value of 1 assigned to those children who scored at or above the 25th percentile and a value of 0 assigned to those who scored below the 25th percentile.

Three CMA tasks were found to be significant predictors of TEMA performance below the 25th percentile (TEMA PR25): (a) Object Counting, (b) One-Set Addition and Subtraction, Objects Hidden, and (c) Shape Recognition. Each of these tasks was positively correlated with TEMA PR25, even in the final model of the math screener that included only these three tasks as predictors. Based on this final logistic regression model, the correct classification rate of 70% (our target classification rate) is reached when the predicted probability cutoff is 0.7. Thus, the math screener for this study included items that measured Object Counting (five items), One-Set Addition and Subtraction (four items), and Shape Recognition (four items) for a total of 11 items. Using the predicted probability cutoff of 0.7, it was determined that a child who received a total correct score of 4 or less on the math screener would be likely to score below the 25th percentile on the TEMA-3. These children were predicted to be at risk for low math achievement in kindergarten and, therefore, were eligible to participate in this tutorial intervention study.

It is important to acknowledge that although incoming academic abilities are good proxies for risk status for school-age children (Compton et al., 2012; Schatschneider, Wagner, & Crawford, 2008), it is not known whether this is also true for very young children with no history of prior schooling. Although children who performed low on the math screener were also low on other math assessments at pretest, the false positive rate for significant risk for math difficulties may be elevated in this and similar studies of pre-K children.

Child Math Assessment (CMA)

The Child Math Assessment (Starkey & Klein, 2012) is a measure of preschool children's informal mathematical knowledge across a broad range of concepts and skills, including number, arithmetic operations, space and geometry, measurement, and patterns. The CMA is sensitive to the development of some math concepts supported by the tutorial-based math intervention, but it was not over-aligned with PKMT (i.e., it does not use the same tasks or materials). The CMA is comprised of nine tasks, with multiple items per task, and the range of task difficulty is appropriate for children from three to five years of age. All tasks on the CMA are administered individually to children in one 20-minute session, and the instrument is available in both English and Spanish. The psychometric properties of the CMA are very good for preschool-aged children. Test–retest reliability over a two-week interval is .91, and internal consistency (stratified coefficient alpha) is .92. Furthermore, with respect to concurrent validity, CMA scores were found to be positively related to TEMA-3 scores ($r = .74$, $p < .01$) among 4- and 5-year-old children.

Test of Early Mathematics Ability, 3rd Edition (TEMA-3)

The TEMA-3 (Ginsburg & Baroody, 2003) is a measure of informal and formal mathematical knowledge in the areas of number and operations. It is a valid measure of early numerical abilities, including number sense, number fact knowledge, arithmetic calculation, and problem solving. The TEMA-3 is designed for use with children aged 3 to 8 years, and it provides norms for performance in this age range. It is administered individually to preschool children in one 20–30 minute session. As reported in the manual, test–retest reliability ranges from .82 to .93, and alternate form reliability is .97. Concurrent validity with other criterion math measures ranges from .54 to .91.

Attention Networks Test (Child-ANT)

The Child-ANT (Rueda, Posner, & Rothbart, 2004) measures alerting (vigilance) and executive aspects of attention in a preschool-friendly version of the standard flanker task, and is used as the measure of attention for the NIH Toolbox for the Assessment of Neurobiological and Behavioral Function–Cognition Battery (Bauer & Zelazo, 2014). The test–retest reliability of this task is .92 and convergent validity with the WPPSI-III Block Design is .60 (Zelazo et al., 2013). In this version, the student had to determine which way the middle fish in a set of three is swimming and catch it with a net using the "Z" or "?" key on a laptop. Before the child completed the task on the computer, an "off computer" practice with cards was used to familiarize the child with the task requirements and several practice trials were administered. The child then practiced on the computer and was provided scaffolding as needed by the examiner depending on the child's responses during practice. After the computerized practice, the child completed the actual test trials, which consisted of four blocks of 17 trials each. Trials were either cued or uncued. On cued tasks, bubbles appeared on the screen before the fish appeared. And trials were either congruent or incongruent. On congruent trials the middle fish and the flanker fish were swimming in the same direction; on the incongruent trials the middle fish was swimming in the opposite direction of the two flanking fish. Examiners were allowed to prompt for attention once per block, but were otherwise discouraged from interacting with the child during the test trials. Accuracy on congruent and incongruent trials and on cued and uncued trials were used in analyses. Response times were also collected, but were not analyzed in the impact analyses because of low levels of accuracy.

Response time analyses are not considered to be meaningful, particularly for preschool children, when accuracy is low (Davidson, Amso, Anderson, & Diamond, 2006).

Kaufman Brief Intelligence Test, Second Edition (KBIT-2)

The KBIT-2 (Kaufman & Kaufman, 2004) Matrices subtest was administered at pretest. This task is used to assess nonverbal cognitive abilities in individuals from 4 to 90 years of age. In this subtest, examinees see a page with one picture at the top and a choice of pictures below. The items involve concrete stimuli (people and objects) as well as more abstract stimuli (designs and symbols). The task is to choose the picture on the bottom that is related to the top picture. Internal consistency reliability is reported to be .86 for Matrices from 4–18 years of age.

Data Collection Procedures: Assessor Training and Child Assessment

Assessors were trained and certified by members of research teams in both states. Assessors received a two-day training workshop on the TEMA-3 and CMA and a one-day training workshop on the attention measure and other cognitive measures. Assessors were certified to a high level of accuracy on a task-by-task basis. Bilingual assessors were certified to deliver assessments in both English and Spanish. For the CMA and TEMA-3, assessors were certified through observation of their testing of pre-K children who were not part of the study. This occurred in situ for California-based assessors and by videotape for Texas-based assessors. Certification on computerized and noncomputerized attention and cognitive measures was conducted during lab-based observations of the delivery of these tasks. Over the first few days of pretesting, supervisors shadowed assessors on all measures as an additional check on their performance in the field. Assessors underwent refresher training and an in-lab recertification process in each state prior to posttesting cohort 1, and prior to pretesting and posttesting cohort 2.

Testing was conducted in several sessions across approximately a four-week period at pretest and during approximately a three-week period at posttest. During pretest, the first session was devoted to the whole-class math screening process. The remaining sessions for students who met the inclusion criteria and who were consented consisted of the math, attention, and cognitive measures. At posttest, the same measures were given in the same order as pretest.

Analysis

We fit three-level regression models (HLM7; Raudenbush, Bryk, & Congdon, 2010) to estimate the effects for math treatment on math outcomes and the effects for attention treatment on attention-related outcomes. Students were nested within classrooms and classrooms were nested in schools. Cases were assigned within classrooms to one of three groups: math intervention (*M only*), combined math and attention intervention ($M + ATT$), and business-as-usual (BaU). For all models, we included pretest scores (group-mean centered) at the student level. We considered three possible effects for math treatment: (a) *M only* versus BaU, (b) $M + ATT$ versus BaU, and (c) *M only* combined with $M + ATT$ (labeled *condition* in Table 3) versus BaU, but only when the *M only* and the $M + ATT$ groups did not differ on math outcomes at posttest. For attention outcomes, we modeled the

Table 3. Pretest and posttest means, standard deviations for math and attention outcomes.

	Pretest			Posttest		
Measures	*M*	*SD*	*N*	*M*	*SD*	*N*
Math Outcomes						
Child Math Assessment						
Full Sample						
M + ATT	0.29	0.11	181	0.61	0.13	175
M only	0.28	0.11	180	0.64	0.15	172
BaU	0.26	0.12	180	0.52	0.15	171
Texas Sample						
M + ATT	0.29	0.12	93	0.61	0.13	88
M only	0.30	0.10	92	0.67	0.14	87
BaU	0.26	0.12	92	0.55	0.15	86
California Sample						
M + ATT	0.28	0.10	88	0.61	0.14	87
M only	0.26	0.12	88	0.60	0.14	85
BaU	0.26	0.12	88	0.49	0.15	85
Test of Early Mathematics Ability (TEMA-3)						
Full Sample						
M + ATT	3.38	2.96	181	13.51	6.00	175
M only	3.17	2.49	180	14.09	6.08	172
BaU	3.28	2.79	180	12.82	6.86	171
Texas Sample						
M + ATT	3.57	3.13	93	14.83	6.25	88
M only	3.62	2.54	92	16.59	6.48	87
BaU	3.38	2.72	92	15.43	7.19	86
California Sample						
M + ATT	3.17	2.78	88	12.18	5.45	87
M only	2.70	2.36	88	11.53	4.37	85
BaU	3.17	2.87	88	10.18	5.37	85
Attention Outcomes						
Cued Trial Accuracy						
Full Sample						
M + ATT	0.69	0.17	179	0.87	0.14	174
M only	0.68	0.18	180	0.83	0.16	171
BaU	0.67	0.18	180	0.83	0.18	171
Texas Sample						
M + ATT	0.69	0.17	92	0.87	0.14	88
M only	0.70	0.17	92	0.87	0.14	86
BaU	0.67	0.18	92	0.83	0.18	86
California Sample						
M + ATT	0.68	0.17	87	0.87	0.13	86
M only	0.66	0.18	88	0.80	0.18	85
BaU	0.66	0.17	88	0.82	0.17	85
Uncued Trial Accuracy						
Full Sample						
M + ATT	0.67	0.17	179	0.85	0.14	174
M only	0.66	0.18	180	0.82	0.17	171
BaU	0.65	0.18	180	0.82	0.17	171
Texas Sample						
M + ATT	0.68	0.18	92	0.85	0.15	88
M only	0.68	0.17	92	0.84	0.16	86
BaU	0.65	0.17	92	0.83	0.17	86
California Sample						
M + ATT	0.65	0.17	87	0.86	0.14	86
M only	0.64	0.18	88	0.79	0.17	85
BaU	0.65	0.18	88	0.81	0.16	85

(*Continued on next page*)

Table 3. (Continued).

Measures	Pretest			Posttest		
	M	*SD*	*N*	*M*	*SD*	*N*
Congruent Trial Accuracy						
Full Sample						
M + ATT	0.84	0.17	179	0.96	0.06	174
M only	0.83	0.18	180	0.94	0.10	171
BaU	0.79	0.19	180	0.93	0.12	171
Texas Sample						
M + ATT	0.83	0.18	92	0.95	0.07	88
M only	0.85	0.16	92	0.94	0.08	86
BaU	0.78	0.20	92	0.92	0.14	86
California Sample						
M + ATT	0.84	0.17	87	0.96	0.05	86
M only	0.82	0.21	88	0.93	0.11	85
BaU	0.81	0.18	88	0.94	0.10	85
Incongruent Trial Accuracy						
Full Sample						
M + ATT	0.51	0.26	179	0.77	0.26	174
M only	0.50	0.27	180	0.71	0.30	171
BaU	0.52	0.24	180	0.72	0.28	171
Texas Sample						
M + ATT	0.53	0.26	92	0.77	0.27	88
M only	0.53	0.26	92	0.76	0.27	86
BaU	0.54	0.24	92	0.75	0.27	86
California Sample						
M + ATT	0.48	0.26	87	0.77	0.25	86
M only	0.47	0.27	88	0.66	0.31	85
BaU	0.50	0.23	88	0.69	0.29	85

effect of $M + ATT$ versus BaU and M *only* combined, but only when the two latter groups did not differ on attention outcomes. We tested variation in treatment effects (random effect) across schools and classrooms via the likelihood ratio tests ($-2\Delta LL$), with degrees of freedom equal to the difference in the number of estimated parameters. The addition of random slopes did not significantly improve model fit for any of the outcomes. Accordingly, we model treatment as a fixed effect.

State (Texas or California) was included as a fixed effect in the level-3 models. We estimated its moderating effect on the relationships of treatments and outcomes by including the appropriate cross-level interaction terms. We used an iterative model-fitting process, fitting fully specified models (i.e., intercept, covariates, treatment effects, and interaction terms), removing nontreatment-related terms that were statistically nonsignificant, and then estimating final models to include statistically significant covariates along with the associated treatment effects for each outcome. We describe the results of these model-building exercises; however, we interpret treatment effects in the context of the fully conditional models. We calculated effect sizes as Hedges's g, using the coefficient corresponding to the relevant parameter as the numerator and the posttest pooled standard deviation as the denominator, per recommendations of the What Works Clearinghouse.

Finally, we used the Benjamini-Hochberg correction (BH; Benjamini & Hochberg, 1995; 2000) to correct for false discovery rates (FDR) associated with multiple comparisons in the context of hierarchical data (Benjamini & Bogomolov, 2011; Yekutieli, 2008). We applied the correction separately to math outcomes and attention outcomes. We calculated q, which is the threshold p value provided by BH-adjustment, for initial models and for final models. Contrasts from the initial models were included in the final models only if they met the adjusted p value (i.e., the q threshold). Decisions about which p values to include in corrections for false discovery are best conceptualized in the context of a study's statistical power, particularly when data are analyzed in multivariate models (Benjamini & Bogomolov, 2011).

The *study* was powered to detect differences between $M + ATT$ and BaU on CMA and TEMA outcomes, the difference between *M only* and BaU on CMA and TEMA outcomes, the difference in the combined $M + ATT/M$ *only* group and BaU on CMA and TEMA outcomes, and *state* differences in CMA and TEMA. The p values associated with these seven effects were used to establish q.

Results

Preliminary Analyses

We evaluated the distributional properties for the outcome variables. None departed significantly from normal. There were no outlying values. Q-Q plots suggested that level-1 residuals were normally distributed and that random effects (using empirical Bayes estimates) were multivariate normal. Assumptions about equal variance along the regression plane were evaluated using scatter plots of level-1 residuals. Errors at different levels of the model were uncorrelated. The CMA reports proportion-correct scores; however, we elected *not* to transform the score distribution because the descriptive values indicated an absence of skewness, ceiling effects, or floor effects. Raw scores were used in analyses of the TEMA-3 outcomes.

Table 3 presents means and standard deviations at pretest and posttest across the three treatment groups for the full sample and for the Texas and California samples. There were no statistically significant group differences at *pretest* on the CMA ($\beta = .02$, $SE = .012$, $p =$ .06 for $M + ATT$ versus BaU; $\beta = .02$, $SE = .01$, $p = .15$ for *M only* versus BaU; $\beta = .02$, $SE = .01$, $p = .08$ for $M + ATT$ versus *M only*), or on the TEMA ($\beta = .02$, $SE = .29$, $p = .95$ for $M + ATT$ versus BaU; $\beta = -.19$, $SE = .29$, $p = .52$ for *M only* versus BaU; $\beta = -.08$, $SE =$ .25, $p = .74$ for $M + ATT$ versus *M only*), suggesting that groups were comparable on math outcomes prior to treatment. Similarly, there were no pretest differences on attention outcomes ($\beta = .02$, $SE = .01$, $p = .36$ for cued trial accuracy; $\beta = .01$, $SE = .02$, $p = .45$ for uncued trial accuracy; $\beta = .03$, $SE = .02$, $p = .08$ for congruent trial accuracy; and $\beta = -.003$, $SE = .02$, $p = .87$ for incongruent trial accuracy). There were no differences in the distribution of gender across conditions (55.2% male and 44.8% female in $M + ATT$; 49.7% male and 50.3% female in *M only*, and 52.5% male and 47.5% female in BaU).

There were significant group differences on *exposure* (operationalized as the difference in student age at pretest and posttest) for the $M + ATT$ versus BaU contrast and for the *M only* versus BaU contrast on the CMA ($\beta = .22$, $p = <0.001$ for $M + ATT$; $\beta = .23$, $p = <0.001$ *M only*) and on the TEMA ($\beta = .17$, $SE = .02$, $p = <0.001$ & $\beta = .17$, $SE = .02$, $p = <0.001$ for *M only* and $M + ATT$, respectively), meaning that the amount of time between pretest and posttest was greater, on average, for students in the $M + ATT$ and in the *M only* groups compared to BaU. Accordingly, we included *exposure* as a level-1 fixed effect in CMA and TEMA models as a control. There were no group differences in *exposure* on the attention outcomes.

Math Outcomes

Child Math Assessment

Prior to estimating treatment effects on CMA, we contrasted the $M + ATT$ and the *M only* groups on math outcomes as described earlier. The groups differed significantly ($\beta = -.03$, $SE = .01$, $p = .02$), precluding the possibility of combining the two as a contrast with *BaU*. Instead, we compared each to *BaU* independently. Thus, the initial models for CMA included the $M + ATT$ and *M only* (i.e., treatment) effects along with pretest scores, *state*, and a *state* by treatment interaction, as described above. Additionally, given the statistically significant group difference for *exposure*, we included its main effect and its interaction with *state*.

Initial CMA Model. In the initial, fully specified model, $M + ATT$ participants outperformed children in the *BaU* at posttest ($\beta = .05$, $SE = .02$, $p = 0.01$), with a moderately sized treatment effect ($g = .36$). The *M only* group also outperformed the *BaU* ($\beta = .10$, $SE = .02$; $p < 0.001$). The effect size (Hedges's *g*) was .67. The coefficient for *state* differed from 0 ($\beta = -.06$, SE $= .02$ $p = 0.01$), suggesting that CMA posttest scores in Texas were higher, on average, than scores in California. The effect for *exposure* ($\beta = .06$, $SE = .02$, $p = 0.01$) was statistically significant; larger intervals between pretest and posttest were associated with higher CMA scores. There were no statistically significant cross-level interaction effects for *state* or for *exposure* ($\beta = .03$, $SE = .03$ $p = 0.31$ for *state* by $M + ATT$; $\beta = -.02$, $SE = .03$ $p = 0.58$ for *state* by *M only*; $\beta = .07$, $SE = .05$ $p = 0.20$ for *state* by *exposure*).

Final CMA Model. The final model, reestimated to include only significant predictors, is summarized in Table 4. The $M + ATT$ effect was .43 (Hedges's *g*) and the *p* value was less than .001. The coefficient for *M only* was .09 ($SE = .01$, $p < 0.001$), and the effect size (Hedges's *g*) was .60. The coefficient for *state* differed significantly from 0 ($\beta = -.06$, $SE = .02$, $p < 0.001$) and the effect for *exposure* ($\beta = .07$, $SE = .02$, $p = 0.01$) was statistically significant. Finally, the cluster-level random effect, r_0, differed statistically from 0 ($p < 0.01$).

Test of Early Mathematics Ability-3

There were no differences in the posttest means for the $Math + ATT$ group and the *M only* group ($\beta = -.74$, $SE = .50$, $p = .14$) on TEMA; the two groups were combined to estimate treatment effects. In a separate model, we contrasted the two treatment groups, $M + ATT$ and *M only*, with *BaU* as described in an earlier section, to allow for comparison with the CMA results and for purposes of the B-H correction.

Initial TEMA Model. We describe the initial noncombined model first. When *BaU* was contrasted with $M + ATT$ and with *M only*, there were no main effects for either treatment ($\beta = -.63$, $SE = .74$, $p = 0.39$ for $M + ATT$ versus *BaU*; $\beta = .86$, $SE = .75$, $p = 0.25$ for *M only* versus *BaU*). There were also no effects for *exposure* ($\beta = .96$, $SE = .92$, $p = 0.30$) when modeled as a predictor of TEMA scores (contrary to the results for the preliminary analysis where exposure was the dependent variable). There were *state* differences ($\beta = -5.711$, $SE = .93$ $p = < 0.001$), indicating that TEMA posttest scores in Texas were higher, on average, than scores in California. Additionally, the *state* by $M + ATT$ cross-level interaction was statistically significant ($\beta = 2.39$, $SE = 1.12$, $p = 0.03$), meaning that the difference

Table 4. Fixed and random effects for math outcome analysis.

		Fixed effects						
		Unconditional model			Conditional model			
Measures	Predictor	Coefficient[a]	*SE*	*p* value	Coefficient[a]	*SE*	*p* value	Effect size
CMA								
	Intercept, γ_{000}	0.58801	0.00955	<0.001	0.16050	0.14575	0.28	
	State, γ_{001}				−0.05815	0.01638	<0.001	
	Pretest, γ_{100}				0.66807	0.04848	<0.001	
	Math + ATT, γ_{200}				0.06149	0.01312	<0.001	.43
	Math-only, γ_{300}				0.09050	0.01320	<0.001	.60
	Exposure, γ_{400}				0.06559	0.02112	0.01	
TEMA								
	Intercept, γ_{000}	13.23847	0.48113	<0.001	20.43586	1.28039	<0.001	
	State, γ_{001}				−5.09582	0.77207	<0.001	
	Pretest, γ_{100}				1.09900	0.08969	<0.001	
	Math + ATT, γ_{200}				−0.30872	0.69643	0.66	.05
	Math-only, γ_{300}				1.52169	0.53016	0.01	.23
	*M + ATT**State, γ_{400}				2.11406	0.91037	0.02	
TEMA								
	Intercept, γ_{000}	13.23847	0.48113	<0.001	19.36626	1.19927	<0.001	
	State, γ_{001}				−4.37959	0.70993	<0.00	
	Condition, γ_{100}				1.13087	0.46239	.02	.18
	Pretest, γ_{200}				1.10136	0.09037	<0.001	
		Random effect						
		Estimate	*p* value	%[b]	Estimate	*p* value	%[b]	
CMA								
	r_0 *(Level 2)*	0.00047	.128	1.99	0.00254	<0.001	16.34	
	e (Level 1)	0.02160		91.68	0.01265		81.40	
	u_{00} *(Level 3)*	0.00149	.003	6.32	0.00035	.185	2.25	
TEMA								
	r_0 *(Level 2)*	1.26371	.10	3.17	2.8146	.00	10.06	
	e (Level 1)	32.9043		82.46	23.6964		84.70	
	u_{00} *(Level 3)*	5.73569	<0.001	14.37	1.46625	.04	5.24	
TEMA								
	r_0 *(Level 2)*	1.26371	.10	3.17	2.6979	.002	9.52	
	e (Level 1)	32.9043		82.46	24.12128		85.13	
	u_{00} *(Level 3)*	5.73569	<0.001	14.37	1.51478	.037	5.35	

Note.[a]Coefficients are unstandardized; [b]percentage of variance explained.

between the *M + ATT* and *BaU* groups in the California sample was greater than the difference between the *M + ATT* and *BaU* groups in the Texas sample. Although Texas children in *M + ATT* group scored higher than California children in the *M + ATT* group at posttest, California children made greater gains, on average, than their Texas counterparts. *State* did not moderate the effect of *M only* on TEMA when contrasted to *BaU* ($\beta = 0.99$, $SE = 1.11$, $p = 0.37$). In the second model, the combined *M + ATT/M only* group did not outperform the *BaU* ($\beta = 0.14$, $SE = .73$, $p = 0.85$). The *state* by *condition* interaction also did not differ from 0 ($\beta = 1.21$, $SE = 1.19$, $p = 0.31$) nor did effect for *exposure* ($\beta = 1.72$, $SE = .91$, $p = 0.06$).

Final TEMA Model. The refit models for TEMA are summarized in the lower half of Table 4. In the trimmed noncombined groups' model, the main effect for *M + ATT* group was not significant when contrasted with *BaU* ($\beta = -0.31$, $SE = 0.70$, $p = 0.66$). However, as before, its interaction with *state* did differ from 0 ($\beta = 2.11$, $SE = 0.91$, $p = 0.02$), suggesting that the treatment effect associated with $M + ATT$ in California was greater than the corresponding treatment effect in Texas ($g_{CA} = 0.38$; $g_{TX} = -0.10$). The *M only* contrast with *BaU* was statistically significant ($\beta = 1.52$, $SE = 53$, $p = 0.01$), as well, suggesting that children in *M only* outperformed *BaU* children in both states. The effect size was .23 (Hedges's *g*).

In the final model for the combined groups (*condition* in Table 4), the contrast with *BaU* was statistically significant ($\beta = 1.13$, $SE = .46$ $p = 0.02$, $g = .18$), with children in the treatment groups scoring higher on the TEMA at posttest than children in the *BaU* group.

Correction for False Discovery Rate on Math Outcomes

The adjusted *p* value for the *initial* math models was .028 ($q = .028$). The adjustment was based on the treatment-related contrasts and the main effect for state on CMA and TEMA. The original power analysis for the study was based on these contrasts. We included main effects in the final model only if they met the $q = .028$ threshold. We modeled treatment-related interactions in the initial models. We included the interaction terms in the refit final models if the fixed-effect for *state* in the initial model met the *q* threshold. In the *final* math models, *q* was .044, meaning that all coefficients with a *p* value of .044 or less were statistically significant after adjusting for false discovery rate. Final model estimates for the primary contrasts were evaluated at the .044 level (i.e., effects described as significant in the foregoing presentation met this standard). We also evaluated the interaction terms at the *q* threshold.

Our approach follows the recommendations of Yekutieli (2008) and Benjamini & Bogomolov (2011) for estimating *q* in the context of multilevel, multivariate data where statistical power is established a priori. Whereas Yekutieli (2008) works in the context of DNA microarray data, which often involves tens of thousands of contrasts, the logic that underlies his "tree of hypotheses" approach applies to analyses conducted on a smaller scale. Benjamini & Bogomolov (2011) describe FDR corrections across a variety of complex analytic contexts, contrasting the strategic use of BH to the "blunt instrument approach" associated with corrections for overall family-wise error (e.g., Bonferroni). In the context of prospective experimental studies, Benjamini & Bogomolov (2011) recommend that FDR corrections be based on the contrasts for which the study was initially powered.

Attention Outcomes

Because accuracy on the ANT was relatively low (for more than one ANT measure and for at least one time point), only accuracy (not response time) data were analyzed in keeping with approaches taken in other studies of preschool children (Davidson et al., 2006; reviewed in Zelazo et al., 2013). Incongruent trials, congruent trials, cued trials, and uncued trials were analyzed as the main measures of impact for attention training. Incongruent and congruent trials collapsed across the "cueing" variable whereas cued and uncued trials collapsed across the "congruency" variable. To the extent that vigilance is improved by attention training, performance on both cued and uncued trials as well as congruent trials would be expected to improve. To the extent that executive attention is improved by attention training, better performance should be observed on incongruent trials in which the direction of the flanker fish conflicts with the direction of the central fish, requiring conflict resolution.

Initial Attention Model

Comparisons of *M only* and *BaU* on attention outcomes were not statistically significant [cued trial accuracy ($\beta = -0.003$, $SE = 0.0174$, $p = .800$), uncued trial accuracy ($\beta = -0.006$, $SE = 0.017$, $p = .724$), congruent trial accuracy ($\beta = 0.005$, $SE = 0.012$, $p = .656$), and incongruent trial accuracy ($\beta = 0.006$, $SE = 0.031$, $p = .984$)]. Accordingly, the two groups were combined and contrasted with *Math + ATT* to evaluate the effects of attention treatment, as described elsewhere. *Condition* represents the effect of *Math + ATT* group compared to the combined *M only* and *BaU* comparison. Finally, the false discovery rate equals the family-wise error rate (FWER) when the number of true null hypotheses (m_0) equals the number of all hypotheses (m) under test (Benjamini & Hochberg, 2000). Correcting for FDR also corrects for FWER. Thus, α in the attention-related contrasts was established at .05.

Final Models for Attention Outcomes

Table 5 summarizes the findings for attention-related outcomes. The main effect of state did not differ statistically from 0 for any attention-related outcomes (Table 5). For Cued Trial Accuracy, children in the *M +ATT* condition outperformed students in the combined *M only* and *BaU* group ($\beta = 0.04$, $SE = 0.01$, $p = .01$). The effect size was 25. The treated group (*Condition*) scored higher at posttest than the comparison on the measure of Uncued Trial Accuracy ($\beta = 0.03$, $SE = 0.01$, $p = .04$; $g = 0.19$), on the test of Congruent Trial Accuracy ($\beta = 0.02$, $SE = 0.01$, $p = .03$, $g = .21$), and on the measure of Incongruent Trial Accuracy ($\beta = 0.05$, $SE = 0.02$, $p = .03$, $g = 0.18$).

Discussion

The current study sought to identify a sample of low-income preschool children who were very low performing in math early in the school year, and tested the efficacy of an intensive math intervention, with or without additional attention training, to improve the math outcomes for these at-risk children. Intensive interventions at higher tiers of instruction have been tested with pre-kindergartners in the area of early literacy (e.g., Lonigan & Phillips, 2016). To our knowledge, this study is the first to test a novel approach to increasing the intensity of intervention for low-performing pre-kindergarten children in mathematics. Our approach was twofold. We tested the effects of a tutorial-based small-group intervention in math (PKMT) that supplemented, but did not supplant, Tier 1 math instruction. In addition, in order to address the learning-related cognitive weaknesses of children with math difficulties, we tested the effects of this supplemental math intervention in combination with training in attention, a domain-general ability that has been strongly associated with math development and disability, and response to intervention (Tannock, 2013).

We hypothesized that children who received their classroom math instruction plus the PKMT intervention (*M only* and *M + ATT* groups) would have higher math scores at posttest than children who received only Tier 1 math instruction (*BaU* group). We found significant effects for children who received the PKMT intervention on the CMA, which is a broad measure of informal mathematical knowledge. In independent contrasts with the *BaU* group, both the *M + ATT* group (ES = .43) and the *M only* group (ES = .60) demonstrated

Table 5. Fixed and random effects for attention outcome analysis.

		Fixed effects						
		Unconditional model			Conditional model			
Measures	Predictor	Coefficient[a]	*SE*	*p* value	Coefficient[a]	*SE*	*p* value	Effect size
Cued								
	Intercept, γ_{000}	0.84444	0.00810	<0.001	0.84522	0.01190	<0.001	
	State, γ_{001}				−0.02551	0.01574	0.11	
	Pretest, γ_{100}				0.41550	0.03808	<0.001	
	Condition, γ_{200}				0.03521	0.01283	0.01	.25
Uncued								
	Intercept, γ_{000}	0.82943	0.00844	<0.001	0.83245	0.01271	<0.001	
	State, γ_{001}				−0.02316	0.01658	0.17	
	Pretest, γ_{100}				0.41424	0.03744	<0.001	
	Condition, γ_{200}				0.02643	0.01284	0.04	.19
Congruent								
	Intercept, γ_{000}	0.93877	0.00527	<0.001	0.92753	.00842	<0.001	
	State, γ_{001}				0.00878	.01075	.42	
	Pretest, γ_{100}				0.22795	.02343	<0.001	
	Condition, γ_{200}				0.01754	.00798	.03	.21
Incongruent								
	Intercept, γ_{000}	0.73020	0.01419	<0.001	0.74222	.02129	<0.001	
	State, γ_{001}				−0.05774	.02630	.03	
	Pretest, γ_{100}				0.39377	.04467	<0.001	
	Condition, γ_{200}				0.05287	.02383	.03	.18

		Random effect					
		Estimate	*p* value	%	Estimate	*p* value	
Cued							
	r_0 *(Level 2)*	0.00107	0.028	3.17	0.00212	<0.001	10.06
	e (Level 1)	0.02423		82.46	0.01847		84.70
	u_{00} *(Level 3)*	0.01226	>.500	14.37	0.00030	>.500	5.24
Uncued							
	r_0 *(Level 2)*	0.00064	0.072	2.52	0.00178	.01	8.63
	e (Level 1)	0.02428		95.59	0.01849		89.67
	u_{00} *(Level 3)*	0.00048	0.294	1.89	0.00035	.339	1.70
Congruent							
	r_0 *(Level 2)*	0.00000	.253	0.00	0.00035	.03	4.48
	e (Level 1)	0.00893		96.44	0.00712		91.17
	u_{00} *(Level 3)*	0.00033	.032	3.56	0.00034	.03	4.35
Incongruent							
	r_0 *(Level 2)*	0.00212	.196	2.74	0.00418	.05	6.19
	e (Level 1)	0.07428		96.13	0.06334		93.81
	u_{00} *(Level 3)*	0.00087	.300	1.13	0.00000	.43	0.00

Note.[a]Coefficients are unstandardized.

greater math knowledge at posttest. There was also a main effect for state on the CMA, indicating that the children in Texas classrooms had higher scores at posttest than the children in California classrooms. The effects of the intervention on the TEMA-3, which measures only numerical knowledge, were significant, but smaller. There was a significant effect for the *M only* group contrasted with the *BaU* group (ES = .23), showing that the *M only* group had higher posttest scores in both Texas and California. The significant effect of the *M + ATT* group, however, varied with state. Although the *M + ATT* group in California scored higher at posttest than the BaU group, this effect was not found in Texas. We discuss state differences in math outcomes in more detail below.

We predicted that attention training would result in higher attention scores for children in the $M + ATT$ group compared to the two groups that did not receive attention training—the *M only* and *BaU* groups. There were significant but small effects of attention training on attention outcomes. Finally, to the extent that attention training facilitates mathematical learning and performance, we asked whether children in the $M + ATT$ group would have higher math scores than those in the *M only* group. There was no evidence that attention training in combination with the PKMT had an additive or synergistic effect on either math outcome.

A positive effect of the math intervention was observed on the CMA, which is a broad measure of early math knowledge. The CMA assesses number, arithmetic, space and geometry, measurement, and patterns and thus was more aligned with the mathematical content of the PKMT intervention than our second measure, the TEMA-3. The effect of the PKMT intervention on the TEMA-3 was smaller, though statistically significant, than on the CMA. We consider several possible explanations for the differences in effect sizes for the two outcome measures. First, the TEMA-3 assesses only number and operations. One implication of the TEMA's more restricted focus is that it would not be able to capture growth in nonnumerical areas of math knowledge (e.g., space and geometry) supported by the PKMT intervention. Thus, the difference in effects across the two outcome measures could be related to the differential alignment of each test with the math tutorial intervention. Second, smaller effects of interventions for struggling learners are not uncommon in more contemporary studies, and these smaller effects have been linked to improvements in the counterfactual or business-as-usual condition in recent studies compared to earlier studies (Lemons, Fuchs, Gilbert, & Fuchs, 2014; Scammacca, Roberts, Vaughn, & Stuebing, 2015). We turn next to issues regarding the nature of the counterfactual and its relevance for interpreting the findings of the current study.

There were main effects of state for the CMA measure, with Texas children scoring higher than children from California. Furthermore, where there were interactions with state, these took the form of larger effects on the TEMA measure for children from California. We interpret these findings in the context of what we know about how the *BaU* condition differed across the two states. The EMCO analyses showed that classroom teachers in Texas delivered about twice as many minutes of math instruction in a given day compared to California classroom teachers. Lemons and his colleagues (Lemons, Fuchs, Gilbert, & Fuchs, 2014) have argued that in order to interpret the effects of interventions in experimental and quasi-experimental studies, it is important to understand that *BaU* is not a static condition, but varies across time and location. In recent years, the use of more research-informed Tier 1 programs is associated with reduced effect sizes in randomized trials.

We believe that this more nuanced view of the effect of the counterfactual is at play in this study. Although we powered the study for possible state differences, we did not hypothesize that there would be state differences for the math outcomes. In the first year of the intervention, the early childhood program in the Texas-based preschools implemented a daily "numeracy instructional block" and provided extensive professional development along with implementation monitoring of pre-kindergarten teachers. The fact that a relatively strong, numeracy-focused Tier 1 program was in place in Texas classrooms is consistent with the finding that the effects on the TEMA (a numeracy-focused measure) were smaller in Texas, especially for the $M + ATT$ group, than in California. Furthermore, the strong Tier 1 program in Texas makes for a stringent test of the value-added benefit of the supplemental PKMT program. In this regard, it is of interest to note that there were still significant effects on the CMA for children receiving the PKMT intervention in Texas. This points to the effectiveness of the intensive math intervention for very low-performing pre-kindergarten children, even in the context of a strong Tier 1 mathematics curriculum.

Although there were significant effects of attention training on attention outcomes, these effects were small and consistent for measures of vigilance and executive attention. These small effects on attention are likely related to the low intensity of the attention intervention, particularly in comparison to the intensity of the math intervention. Our attention training sessions were relatively short and once per week because our piloting showed that most children were unable to tolerate longer sessions and they were also unable to sustain focus when shorter attention-training sessions were attempted on the same day as the math sessions. Difficulties with task engagement and motivation are common issues in cognitive training studies with young children and may account for the finding that cognitive training studies with younger children are of shorter duration and lower intensity compared to those with older children (Wass et al., 2012).

Although the attention-training programs were adaptive at the child level, we did not allow for varying amounts of attention-training time between children or for increases in the length of attention training over time. Whether more intensive attention training would be associated with increases in attention scores and transfer to mathematical learning cannot be determined from the current impact analyses; however, we are investigating whether achieving higher levels on the attention games is associated with stronger effects on attention outcomes, and if so, whether these gains in attention are associated with math outcomes. This issue of intervention intensity is discussed in further detail below.

Another potential concern was the use of one measure to assess attention outcomes. Flanker tasks assess vigilance and executive attention in the same paradigm, which is why the Child-ANT was used in this study. Typically, however, this paradigm yields both accuracy and response time scores, each of which can be used in analyses. In this case, however, because of the young age of the children and low accuracy, we could not use response time measures as another indicator of attention abilities.

One of the strengths of the study design is that we controlled for general attention factors and amount of adult attention given to children in the *M only* and $M + ATT$. The tutor-supervised sessions for children who listened to books on tape were of the same length and frequency as the tutor-supervised attention-training sessions. Furthermore, a simple behavioral system that rewarded “paying attention” was provided to both groups. It is possible that working on a computer once a week had a general facilitative effect for performance on other computerized tasks. However, we did not assess general use of computers in the children’s classrooms as part of the counterfactual.

With respect to the low intensity of attention training in this trial, it is interesting to note that typical attention interventions with young children tend to have relatively low amounts of training time, ranging from times similar to those used in our study (i.e., 120 minutes; Rueda et al., 2005) to somewhat higher amounts of training (e.g., 400 minutes; Rueda et al., 2012). This contrast in intensity of cognitive versus academic interventions may be important given that dosage and session duration are associated with transfer effects of cognitive training (Schwaighofer, Fischer, & Bühner, 2015). We would argue that most interventions that contain a cognitive training component, including our study, have been remarkably nonintensive with respect to dosage and duration of intervention. There is also relatively little research to guide the implementation of nondosage-related features of cognitive training procedures. For example, it is not clear whether the *timing* of cognitive training matters; that is, it is not clear whether training attention before instituting math interventions versus concurrently with the math intervention as was done in this study might be more effective. Whether the type of attention training matters is also largely unexplored; for example, is training in both alerting and executive attention required to produce changes in these two attention systems that have been related to mathematical cognition (Ashkenazi & Henik, 2012; Rueda et al., 2005, 2012)? In sum, there are several aspects of attention training and its effect on both attention and math that are currently unknown, but which deserve further study.

Given the small effects of attention training on attention, we might have expected no effect of attention training on mathematics. In fact, the $M + ATT$ group in Texas had significantly lower math scores at posttest on the CMA and the TEMA than did their peers in the *M only* group (see Table 3). We have no explanation for this counterintuitive finding given that the children in the *M only* and $M + ATT$ groups were tutored together and also in light of the fact that attention training occurred on a separate day from the mathematics intervention. Potential sources of variability between *M only* and $M + ATT$ groups that we have investigated include nonverbal IQ, absenteeism, and other measures of dosage of mathematics instruction. None of these factors appear to explain the differences in math outcomes between the two treatment groups in Texas.

A finding from this study that is of interest when thinking about what intensity of intervention means for very low-performing preschool children is the main effect of state on the CMA outcome measure. Research-based recommendations for increasing intensity of intervention for children with more entrenched learning difficulties include reducing group size and increasing time in intervention (Vaughn et al., 2010). One way to significantly increase time spent in intervention is to leverage the potential daily multiple opportunities for instruction, including research-informed Tier 1 math instruction plus research-validated higher tiers of supplemental small-group intervention (Fuchs et al., 2014). Our findings suggest that prekindergarten children who were low performing in math benefited from the use of intensive interventions in Texas classrooms that began with the core numeracy curriculum and added higher tiers of instruction provided by the PKMT intervention. However, it is worth noting that a subgroup of children at both sites showed relatively low response to their combined core curriculum and supplemental math interventions. Eighteen percent of children in Texas and 30% of children in California were below the 10th percentile on the TEMA-3 at posttest. Similar low performance in kindergarten is a strong predictor of future learning disabilities in mathematics five years later (Morgan et al., 2009).

In conclusion, this study is one of the first to test approaches to increasing intensity of interventions in mathematics for prekindergarten children. The math-specific aspects of the intervention were associated with effects on math outcomes. Effects of attention training on attention were small and did not transfer to mathematics. Although many children in the intervention conditions made considerable gains in mathematical knowledge over the prekindergarten year, it is clear that there is also a subgroup of children who did not make sufficient gains to prepare them for mathematics instruction in kindergarten. Understanding the predictors of low response in this group of children will be important for finding ways to further intensify and tailor early math interventions.

Acknowledgments

The authors wish to acknowledge the invaluable contributions of the staff and students from the Houston Independent School District, San Jose Unified School District, Santa Clara County Office of Education, and West Contra Costa Unified School District.

Funding

The research reported here was supported by grant R342A110270 from the Institute of Education Sciences (NCSER), U.S. Department of Education to the University of Texas at Austin. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education.

ARTICLE HISTORY

Received 22 March 2016
Revised 15 May 2016
Accepted 16 May 2016

EDITORS

This article was reviewed and accepted under the editorship of Carol McDonald Connor and Spyros Konstantopoulos.

References

Ashkenazi, S., & Henik, A. (2012). Does attentional training improve numerical processing in developmental dyscalculia? *Neuropsychology*, *26*(1), 45–56.

Bauer, P. J., & Zelazo, P. D. (2014). The National Institutes of Health toolbox for the assessment of neurological and behavioral function: A tool for developmental science. *Child Development Perspectives*, *8*(3), 119–124.

Benjamini, Y., & Bogomolov, M. (2011). *Adjusting for selection bias in testing multiple families of hypotheses*. Retrieved from https://arxiv.org/pdf/1106.3670.pdf

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289–300.

Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, *25*(1), 60–83.

Bisanz, J., Sherman, J. L., Rasmussen, C., & Ho, E. (2005). Development of arithmetic skills and knowledge in preschool children. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 143–162). New York, NY: Psychology Press.

Blair, C., Ursache, A., Greenberg, M., & Vernon-Feagans, L. (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter–word knowledge in the early elementary grades. *Developmental Psychology*, *51*(4), 459–472.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *Elementary School Journal*, *108*, 115–130.

Bryant, D. P., Bryant, B. R., Gersten, R., Scammacca, N., & Chavez, M. M. (2008). Mathematics intervention for first- and second-grade students with mathematics difficulties: The effects of Tier 2 intervention delivered as booster lessons. *Remedial and Special Education*, *29*, 20–32.

Casey, B., Kersh, J. E., & Young, J. M. (2004). Storytelling sagas: An effective medium for teaching early childhood mathematics. *Early Childhood Research Quarterly*, *19*, 167–172.

Cirino, P. T., Fletcher, J. M., Ewing-Cobbs, L., Barnes, M. A., & Fuchs, L. S. (2007). Cognitive arithmetic differences in learning difficulty groups and the role of behavioral inattention. *Learning Disabilities Research & Practice*, *22*, 25–35.

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review*, *28*(4), 415–427.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, *42*, 127–166.

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., … Bouton, B. (2012). Accelerating chronically unresponsive children to Tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities*, *45*(3), 204–216.

Copley, J. V. (2004). The early childhood collaborative: A professional development model to communicate and implement the standards. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 401–414). Mahwah, NJ: Lawrence Erlbaum Associates.

Copley, J. V., & Padron, Y. (1999). *Preparing teachers of young learners: Professional development of early childhood teachers in mathematics and science*. Retrieved from http://www.project2061.org/publications/earlychild/online/fostering/copleyp.htm

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P, … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446.

Farran, D. C., Silveri, B., & Culp, A. (1991). Public preschools and the disadvantaged. In W. Damon (Series Ed.), L. Rescorla, M. C. Hyson, & K. Hirsh-Pasek (Vol. Eds.), *New directions for child development No. 53: Academic instruction in early childhood: Challenge or pressure?* (pp. 65–73). San Francisco, CA: Jossey-Bass.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. New York, NY: The Guilford Press.

Fuchs, D., & Fuchs, L. S. (2014). Rethinking service delivery for students with significant learning problems: Developing and implementing intensive instruction. *Remedial and Special Education*, *36*(2), 105–111.

Fuchs, D., Fuchs, L. S., & Vaughn, S. (2014). What is intensive instruction and why is it important? *Teaching Exceptional Children*, *46*(4), 13–18.

Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., … Changas, P. (2014). Does working memory moderate the effects of fraction intervention? An aptitude–treatment interaction. *Journal of Educational Psychology*, *106*(2), 499–514.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 3–60.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, *79*(3), 1202–1242.

Ginsburg, H. P., & Baroody, A. (2003). *Test of early mathematics ability* (3rd ed.). Austin, TX: Pro-Ed.

Ginsburg, H. P., Klein, A., & Starkey, P. (1998). The development of children's mathematical thinking: Connecting research with practice. In W. Damon, I. Siegel, & A. Renninger (Eds.), *Handbook of child psychology* (Vol. 4, 5th ed., pp. 401–476). New York, NY: John Wiley.

Ginsburg, H. P., & Russell, R. L. (1981). Social class and racial influences on early mathematical thinking. *Monographs of the Society for Research in Child Development*, *46*(6), 1–69.

Greven, C. U., Kovas, Y., Willcutt, E. G., Petrill, S. A., & Plomin, R. (2014). Evidence for shared genetic risk between ADHD symptoms and reduced mathematics ability: A twin study. *Journal of Child Psychology and Psychiatry*, *55*(1), 39–48.

Griffin, S. (2004). Building number sense with Number Worlds: A mathematics program for young children. *Early Childhood Research Quarterly*, *19*, 173–180.

Griffin, S., Case, R., & Siegler, R. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at-risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 24–49). Cambridge, MA: MIT Press.

Hanich, L., & Jordan, N. (2001). Performance across different areas of mathematical cognition in children with learning disabilities. *Journal of Educational Psychology*, *93*, 615–626.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research*, *85*(4), 512–552.

Jordan, N. C., Huttenlocher, J., & Levine, S. C. (1992). Differential calculation abilities in young children from middle- and low-income families. *Developmental Psychology*, *28*, 644–653.

Jordan, N. C., Kaplan, D., Oláh, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, *77*(1), 153–175.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850–867.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test–Second Edition (KBIT-2)*. Circle Pines, MN: American Guidance Service.

Klein, A., & Starkey, P. (2004). Fostering preschool children's mathematical knowledge: Findings from the Berkeley Math Readiness Project. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 343–360). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Kroesbergen, E. H., van't Noordende, J. E., & Kolkman, M. E. (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Child Neuropsychology*, *20*(1), 23–37.

Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, *43*(5), 242–252.

Lewis Presser, A., Clements, M., Ginsburg, H., & Ertle, B. (2015). Big Math for Little Kids: The effectiveness of a preschool and kindergarten mathematics curriculum. *Early Education and Development*, *26*(3), 399–426.

Lonigan, C. J., & Phillips, B. M. (2016). Response to instruction in preschool: Results of two randomized studies with children at significant risk of reading difficulties. *Journal of Educational Psychology*, *108*(1), 114–129.

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childhood Research Quarterly*, *28*(2), 314–324.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291.

Miller, A. C., Fuchs, D., Fuchs, L. S., Compton, D., Kearns, D., Zhang, W., … Kirchner, D. P. (2014). Behavioral attention: A longitudinal study of whether and how it influences the development of word reading and reading comprehension among at-risk readers. *Journal of Research on Educational Effectiveness*, *7*(3), 232–249.

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, *42*, 306–321.

National Association for the Education of Young Children and National Council of Teachers of Mathematics. (2002). *Early childhood mathematics: Promoting good beginnings*. Washington, DC: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics*. Reston, VA: Author.

Peng, P., & Miller, A. C. (2016). Does attention training work? A selective meta-analysis to explore the effects of attention training and moderators. *Learning and Individual Differences*, *45*, 77–87.

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, *58*, 1–23.

Powell, S. R., Fuchs, L. S., Cirino, P. T., Fuchs, D., Compton, D. L., & Changas, P. C. (2015). Effects of a multitier support system on calculation, word problem, and prealgebraic performance among at-risk learners. *Exceptional Children*, *81*(4), 443–470.

Rabiner, D. L., & Malone, P. S. (2004). The impact of tutoring on early reading achievement for children with and without attention problems. *Journal of Abnormal Child Psychology*, *32*(3), 273–284.

Raghubar, K. P., Cirino, P., Barnes, M. A., Ewing-Cobbs, L., Fletcher, J., & Fuchs, L. (2009). Errors in multi-digit arithmetic and behavioral inattention in children with math difficulties. *Journal of Learning Disabilities*, *42*, 356–371.

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.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2010). *HLM 7.00 for Windows*. Chicago, IL: Scientific Software International.

Rueda, M. R., Checa, P., & Cómbita, L. M. (2012). Enhanced efficiency of the executive attention network after training in preschool children: Immediate changes and effects after two months. *Developmental Cognitive Neuroscience*, *2*, 192–204.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2004). Attentional control and self-regulation. *Handbook of Self-Regulation: Research, Theory, and Applications*, *2*, 284–299.

Rueda, M. R., Rothbart, M., McCandliss, B., Saccomanno, L., & Posner, M. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences USA*, *102*, 14931–14936.

Saxe, G. B., Guberman, S. R., & Gearhart, M. (1987). Social and developmental processes in children's understanding of number. *Monographs of the Society for Research in Child Development*, *52*, 2.

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in Grades 4–12: 1980–2011. *Journal of Learning Disabilities*, *48*(4), 369–390.

Schatschneider, C., Wagner, R. K., & Crawford, E. C. (2008). The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning and Individual Differences*, *18*(3), 308–315.

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50*(2), 138–166.

Shalev, R., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine and Child Neurology*, *47*, 121–125.

Sophian, C. (2004). Mathematics for the future: Developing a Head Start curriculum to support mathematics learning. *Early Childhood Research Quarterly*, *19*, 59–81.

Starkey, P., & Klein, A. (1992). Economic and cultural influence on early mathematical development. In F. L. Parker, R. Robinson, S. Sombrano, C. Piotrowski, J. Hagen, S. Randoph, & A. Baker (Eds.), *New directions in child and family research: Shaping Head Start in the 90s* (pp. 440–449). New York, NY: National Council of Jewish Women.

Starkey, P., & Klein, A. (2008). Sociocultural influences on young children's mathematical knowledge. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 253–276). Charlotte, NC: Information Age Publishing.

Starkey, P., & Klein, A. (2012). *Scaling up the implementation of a pre-kindergarten mathematics intervention in public preschool programs* (Final Report: IES Grant R305K050004). Washington, DC: U.S. Department of Education.

Starkey, P., Klein, A., DeFlorio, L., & Swank, P. (2016). *Scaling up a pre-K mathematics intervention in public preschool programs*. Manuscript submitted for publication.

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, *19*, 99–120.

Tannock, R. (2013). Rethinking ADHD and LD in DSM-5 proposed changes in diagnostic criteria. *Journal of Learning Disabilities*, *46*(1), 5–25.

Vaughn, S., Denton, C. A., & Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools*, *47*(5), 432–444.

Wass, S. V. (2015). Applying cognitive training to target executive functions during early development. *Child Neuropsychology*, *21*(2), 150–166.

Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory—Is younger, better? *Developmental Review*, *32*(4), 360–387.
Yekutieli, D. (2008). Hierarchical false discovery rate-controlling methodology. *Journal of the American Statistical Association*, *103*(481), 309–316.
Zelazo, P. D., Anderson, J. E., Richler, J., Wallner−Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH toolbox cognition battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, *78*(4), 16–33.