Scaling Up the Implementation of a Pre-Kindergarten Mathematics Intervention in Public Preschool Programs

Final Report for IES Grant R305K050004 (2012)

Prentice Starkey and Alice Klein

WestEd

Abstract

A socioeconomic status (SES) gap in mathematical knowledge emerges early and widens prior to school entry. To address this gap, a curricular intervention, *Pre-K Mathematics*, was developed and found to be effective in prior efficacy research. In the present project, the next step was taken in evaluating this intervention. Specifically, an effectiveness study was conducted to determine the degree to which the intervention improves pre-kindergarten (4-year-old) children's mathematical knowledge when implemented by local program staff in multiple settings that serve a heterogeneous population of low-SES families. In contrast with the prior efficacy study, the effectiveness study required that all teachers, rather than volunteer teachers, in their public preschool programs be available for random selection and random assignment. It also used curriculum coaches who were either members of the participating school districts or Head Start programs' permanent training staff or independent contractors, depending on the way a program a routinely supported teacher learning for its in-service teaching staff. Participating programs included publicly funded Head Start and state preschool programs serving low-income, ethnically/racially diverse, urban families in California and low-income, predominantly White, rural families in Kentucky and Indiana. A trainer-of-trainers model was used (1) to train curriculum coaches to support teachers' implementation and (2) to train teachers to implement Pre-K Mathematics with adequate fidelity. A two-condition (treatment and control) RCT was conducted, with clusters of pre-kindergarten classrooms as the unit of randomization. Treatment teachers implemented Pre-K Mathematics and control teachers continued their usual classroom practices. Children were assessed at pretest, posttest, and kindergarten follow up using the Child Math Assessment (CMA) and the Test of Early Mathematics Ability, 3rd Edition (TEMA-3). Classroom observations were made to measure the nature and amount of math support provided by treatment and control teachers during the school year. Coaches supported implementation and teachers implemented with adequate to high levels of fidelity. Multi-level analyses revealed that treatment children made significantly greater gains in mathematical knowledge than control children during the pre-kindergarten year as measured by the CMA (*ES* = 0.83) and TEMA-3 (*ES* = 0.45). A multilevel mediation analysis found evidence that time spent in mathematically focused small-group activities had a significant indirect effect on children's math outcomes. Thus, this effectiveness study found that the Pre-K Mathematics intervention had a significant positive effect on low-SES children's mathematical knowledge. An implication of this finding is that early mathematics intervention is a promising educational strategy for reducing the SES gap in mathematical knowledge.

Additional materials: 73 references, and 1 appendix containing 19 tables and 20 figures

Mathematics Achievement and Development of Children from Low-Income Families

All citizens need a broad range of basic mathematical understanding to make informed decisions in their jobs, households, and communities; in addition, careers require an increasing level of proficiency in mathematics (Glenn Commission, 2000; Kilpatrick, Swafford, & Findell, 2001; U.S. Dept. of Labor, Bureau of Labor Statistics, 2000). Yet since the 1970s, a series of assessments of American students' performance has revealed an overall level of mathematical proficiency well below what is desired and needed (Kilpatrick et al., 2001; Mullis et al., 1997; Mullis et al., 2000). In recent years, the National Council of Teachers of Mathematics (NCTM, 2000) and others have addressed these challenges with new standards to improve mathematics education, and progress has been made at the elementary and middle school levels - especially in schools that have instituted reforms (e.g., Clements, Sarama, & DiBiase, 2004; Fuson, Carroll, & Drueck, 2000; IEA, 2001; Riordan & Noyce, 2001). However, children who live in poverty and who are members of linguistic and ethnic minority groups demonstrate significantly lower levels of achievement (Bowman, Donovan, & Burns, 2001; Denton & West, 2002; Natriello, McDill, & Pallas, 1990).

These achievement differences have roots in early childhood. Children from different sociocultural backgrounds enter elementary school at different levels of readiness for a standards-based mathematics curriculum (Clements, Sarama, & DiBiase, 2004; Starkey, Klein, & Wakeley, 2004; West, Denton, & Germino-Hausken, 2000). Low-income kindergarten (K) children have been found to possess less extensive mathematical knowledge than their middle-income peers (Denton & West, 2002; Griffin, Case, & Siegler, 1994; Jordan, Huttenlocher, & Levine, l992). The same has been found for low-income preschool children (e.g., Ginsburg & Russell, l98l; Hughes, 1986; Jordan, Huttenlocher, & Levine, l994; Saxe, Guberman, & Gearhart, l987; Starkey & Klein, l992). A recent cross-cultural study of early mathematical development in China, Japan, and the United States found that a socioeconomic status- (SES-) related gap in early mathematical knowledge was present by 3 years of age in all three countries (Starkey, 2003; Starkey & Klein, 2008). But, critically, while the gap narrows over the preschool years for children in China (see Figure 1) and Japan where preschools provide relatively rich math instruction, it widens for American children (see Figure 2). These early SES-related differences in mathematical development are conceptually broad, encompassing informal knowledge of number, arithmetic, patterns, space/geometry, and measurement (Starkey, Klein, & Wakeley, 2004). Furthermore, they are persistent. SES-related differences in math achievement identified at the outset of elementary school persist and become more pronounced with time (e.g., Entwistle & Alexander, 1990; Rathbun & West, 2004). The consequences of these early differences for children’s school achievement are underlined by findings from the Early Childhood Longitudinal Study (ECLS) for a large national sample of children who have been followed since beginning K (Rathbun & West, 2004). Assessments of their math and reading knowledge at end of third grade found that the achievement gap identified at K entry between children with family risk factors, including low-income, and their more advantaged peers had widened over the first four years of school.

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. Low-income children receive less support for mathematical development both at home and in preschool. Parents' mathematics beliefs and practices vary with socioeconomic status. Middle-class parents tend to believe the home environment plays a role as great or greater than preschool in preparing young children for school mathematics, whereas more low-income parents believe the preschool environment plays the greater role (Starkey, 2003; Starkey & Klein, 2008; Wakeley, 2002). Low-income mothers expect that preschool teachers will provide instruction in numeracy skills to prepare their children for school (Holloway, Rambaud, Fuller, and Eggers-Pierola, l995). With respect to practices, the frequency with which 4- to 6-year-old children engage in number-related activities at home is positively correlated with the extent of their developed numerical knowledge (Blevins-Knabe & Musun-Miller, 1996). Middle-class parents, in comparison to low-income parents, provide mathematics support that is more frequent, mathematically broader, accompanied by scaffolding, and richer in mathematical language (Deflorio, 2011; Hart & Risley, 1995; Saxe et al., l987). In contrast, many low-income parents provide a comparatively narrow base of support (DeFlorio, 2011). This variation in young children's home learning environments is a likely source of the SES-related differences in early mathematical knowledge.

Children from low-income families also receive less support for mathematical development in their preschool classrooms. Many preschool programs provide a paucity of developmentally appropriate supports for early mathematical development (Bryant, Burchinal, Lau, & Sparling, l994; Graham, Nash, & Paul, 1997). Prekindergarten teachers often use no systematic mathematics curriculum, receive little or no training in early childhood mathematics, are unfamiliar with the mathematics curriculum taught in local elementary schools (Starkey & Klein, 2003), and know little about mathematics standards (Copley, 2004; Copley & Padron, 1999). American preschool programs provide less support than Chinese and Japanese preschools, and within the US, public preschools for lower SES children provide less mathematics support than private preschools enrolling higher SES children (Starkey & Klein, 2008). Many public preschool teachers report that they did not know how to support children's developing numerical abilities (Farran, Silveri, & Culp, 1991) and are providing children with an uneven base of support for mathematical development (Starkey & Klein, 2003, 2008). Indeed, even children enrolled in Head Start make minimal improvement in addition and subtraction knowledge over the prekindergarten year (Zill, et al., 2001).

In summary, both the home and classroom learning environments of young American children tend to be less rich in mathematical support for economically disadvantaged children than for their mid0dle-class peers. Consequently, an SES-related gap in mathematical knowledge appears early and widens during early childhood. In contrast, the SES gap narrows over the preschool years in nations, such as China, that provide greater mathematical support in preschool classrooms on a large scale. Thus, mathematics support varies across contexts in early childhood and is related to children's early mathematical development.

Potential for Early Intervention to Reduce this Gap

The authors developed a prekindergarten math intervention (Klein & Starkey, 2004) and evaluated it in a series of intervention studies (e.g., Klein et al., 2008). These studies indicate that early intervention has the potential to reduce the SES-related gap in children's early mathematical knowledge.

Home math activities. In a randomized controlled trial (Starkey & Klein, 2000), families of Head Start children were randomly assigned to a treatment group, who attended a series of classes to learn home math activities developed for parent-child dyads, or a control group who did not attend the classes. The same results were obtained in two Head Start programs, one serving African-American families and the other serving Hispanic families. Treatment children exhibited significantly greater gains in mathematical knowledge from pretest to posttest than control children. This demonstrated that home math activities provided to parents were effective at enhancing low-income children's mathematical development.

Classroom mathematics curriculum. In quasi-experimental study (Starkey, Klein, & Wakeley, 2004), classroom mathematics activities were evaluated. A group of prekindergarten teachers, half in public programs serving low-income children and half in private programs serving middle-income children, volunteered to participate in the project. Teachers were taught to conduct a set of math activities in their classrooms with children in small groups. The impact of the curriculum on children's mathematical knowledge was measured by the Child Math Assessment (CMA) (see Psychometric Study, below, for a description). The CMA scores of the children in these teachers' classrooms (treatment group) were compared with the scores of children in the same teachers' classrooms the year before the curriculum was implemented (comparison group). Analyses revealed that the treatment children obtained significantly higher CMA scores at the end of the prekindergarten year than the comparison children, and this positive effect was found for both low-income and middle-income children.

Randomized controlled efficacy trial at the classroom level. Next, home and classroom mathematics activities were combined in a single curriculum, published as *Pre-K Mathematics* (Klein, Starkey, & Ramirez, 2002), and evaluated in the IES-funded national Preschool Curriculum Evaluation Research (PCER) program. Classrooms of teachers in Head Start and state-funded preschool programs in California and New York, who volunteered to participate, were randomly assigned to treatment and business-as-usual control conditions. Teachers learned to implement the intervention with fidelity through workshops and on-site facilitation provided by project-supplied trainers. Children's math knowledge was assessed using the CMA/CMA-K in fall and spring of PK and in spring of K. The intervention was found to be effective, with effect sizes at the end of the PK year conservatively calculated by the What Works Clearninghouse to be .55 for cohort 1 (a 62% increase in math knowledge for treatment children relative to control children) and .70 for cohort 2, (a 79% increase in math knowledge for treatment relative to control children) (Klein, et al., 2008). In the K follow-up year, treatment children's math scores were still significantly higher than control children's scores with an effect size of .242. The WWC has assigned this intervention the highest rating of effectiveness (++) with the greatest extent of evidence (medium to large) for early math interventions. In summary, *Pre-K Mathematics* is an effective intervention that reduces the SES-related gap in early mathematical knowledge. This made the intervention a good choice for a scale-up study to determine whether it is effective when implemented at a programwide level of scale, which is customary for curricular adoptions.

Rationale for evaluating Pre-K Mathematics when implemented at scale. Some early childhood programs have proved initially effective when implemented on a small scale, but have produced attenuated effects, at best, when implemented on a large scale (Farran, 1990, p. 508; Gilliam, Ripple, Zigler, & Leiter, 2000). Attenuation of effects has resulted largely from wide variation in fidelity to the original treatment model, for instance, changing the number of lessons or the amount of time devoted to each (Elias, Zins, Gracyk, & Weissberg, 2003; Huberman & Miles, 1984) or altering the content (Borman, Hewes, Overman, & Brown, 2002; Klinger, Ahwee, Pilonieta, & Menendez, 2003). 0

Such findings are informative and suggest the need for scaled-up interventions to be organized in a manner that facilitates sustained implementation with adequate fidelity to the original treatment model. Consequently, for the current project we proposed to scale-up implementation of *Pre-K Mathematics* to a level of scale that is customary for public preschool curriculum adoption: Head Start programs and school-district operated state pre-k programs. We proposed to study both the quality of implementation by programs under conditions of routine educational practice and the effects of implementation on the mathematical knowledge of children enrolled in these programs.

Practical importance of the project. In the absence of an early intervention, low-income children will continue to fall behind in mathematical development, and the SES-related gap in school readiness will persist. The information produced by our project can help policy makers and state-level or national programs make more informed decisions about resources, such as funds for effective curricula and adequate professional development, and operating procedures, such as internal checks on fidelity drift, that local programs will need in order to implement effective mathematics support for low-income children. Enhancing children’s readiness for school mathematics on a large scale, therefore, is educationally meaningful and has practical and policy significance.

Theoretical Foundation of the Intervention.

The early development of mathematical cognition. We assume that the primary conceptual foundations of children’s early mathematical knowledge are the cognitive domains of number and space. These domains are partly structured during infancy (Geary, 1994; Starkey & Cooper, 1980; Starkey, Spelke, & Gelman, 1990). The constraints imposed by this partial structuring enable children to attend to and assimilate mathematically relevant inputs from the environment (Gelman & Williams, 1998). Children first develop informal mathematical knowledge—knowledge that depends on the presence or mental representation of sets of concrete objects (Piaget, 1952). This knowledge develops considerably during the first three years of life (Baroody, 2004; Starkey, 1992), and children often have several mathematical competencies when they enroll in preschool at age three (Bisanz, Sherman, Rasmussen, & Ho, 2005; Ginsburg, Klein, & Starkey, 1998). The extent of children’s knowledge at the beginning of preschool depends on their developmental history, especially the mathematically relevant inputs they received in the first three years of life. Informal mathematical knowledge continues to develop during the preschool years and beyond, and research has identified some of the developmental sequences that occur (e.g., Baroody, 2004; Ginsburg, et al., 1998; Sophian, 1996). Informal mathematical knowledge serves as a conceptual foundation for the acquisition of formal mathematical knowledge—the ability to use abstract numerical notation such as the written numerals (1, 2, 3, etc.) and arithmetic operation signs (+, –, etc.). The transition to formal mathematical knowledge begins at age 4–6 years, depending on children's culture and socioeconomic status (Ginsburg, Klein, & Starkey, 1998).

Theory of learning environments embodied in the intervention. Children's early mathematical knowledge is constrained by a developmental niche (Super & Harkness, 1996), comprised primarily of the home and school learning environments. The mathematical support provided in children's niches partly determine the foundation of informal mathematical knowledge they develop. Mathematical knowledge develops primarily in, or as a consequence of, social activity settings (Vygotsky, 1978)—specifically, settings comprised of children actively participating in concrete mathematics activities with teachers (or parents) who scaffold their learning (Ball & Cohen, 1999). Therefore, math instruction is most effective when teachers possess (a) knowledge of mathematical content, (b) knowledge of milestones in early mathematical development, and (c) knowledge of how curriculum activities can be sequenced to coincide with known sequences in early mathematical development.

Theory of change. A causal model is used to explain the causal relation between the mathematics intervention and children's mathematical knowledge. The main elements are (1) the intervention, which contains the active ingredients that can cause change in children's mathematical knowledge, (2) professional development (PD) support in mathematics for teachers, (3) teacher outcomes produced by the PD, (4) mediation, modeled as aspects of adult-guided mathematical activities that engage children's mathematical cognition, (5) moderation variables, which are variables at multiple levels that may moderate effects of the intervention, and (6) child outcomes, which are changes in children's mathematical knowledge that result from the intervention.

The active ingredients in the intervention are modeled as the mathematics content from the *Pre-K Mathematics* curriculum. Intensive and frequent PD will be the primary means through which teachers become able to deliver the curriculum with both fidelity and understanding (cf., Shulman, 2000). The in-depth, domain-specific PD support that teachers will receive – math focused workshops and on-site training aligned with the mathematics curricula - will ensure that they (a) learn the essential mathematics content comprising the scope and sequence of the math curricula, (b) learn to implement with fidelity, including delivery of that content, and (c) are able to support child engagement and learning of mathematics through explicit, teacher-guided instruction. It is expected that teaching essential mathematics content through effective delivery techniques in school classroom settings will change the nature of teaching and learning opportunities for children. Thus, it is predicted that the mathematics experiences of children will be different in treatment classrooms than in control classrooms, and it is expected that the amount and nature of instructional interactions between teachers and children will be different in treatment and control classrooms. For example, treatment children will spend more time than control children engaged in developmentally sensitive, teacher-scaffolded small-group mathematics activities. Furthermore, the impact of the intervention in treatment classrooms may be moderated by variables at multiple levels.

It is also expected that implementation of the intervention as intended will have a positive and direct causal effect on children's mathematical knowledge. The mathematical content of *Pre-K Mathematics* activities that children engage in, with support from teachers and parents, will influence mathematical problem solving and conceptual access in young children. These mental activities will lead to change in children's mathematical knowledge.

Studies Comprising this Scale-Up Project

This project was comprised of three studies. First, the Main Study was conducted to evaluate the effectiveness of *Pre-K Mathematics* when implemented at scale. Second, a Sustainability Study was conducted to determine whether programs could sustain an effective implementation of *Pre-K Mathematics* beyond the initial year in which intensive professional development was provided. Each study is reported below. Third, a Psychometric Study was conducted to develop and document the psychometric properties of the Child Math Assessment (CMA). This instrument was used, along with the TEMA-3 (Ginsburg & Baroody, 2003), to assess the impacts of the intervention on children's mathematical knowledge in the Main Study and Sustainability Study.

Main Study

Hypotheses and Research Questions

Testing the scale-up of a mathematics curriculum, *Pre-K Mathematics* (Klein & Starkey, 2004), already shown to have beneficial effects involves a shift from a focus on the effectiveness of the curriculum per se to the effectiveness of the proposed process by which preschools and Head Start centers can be assisted to implement the curriculum well enough to attain its benefits. The level of scale used in this project – school district pre-kindergarten programs and grantee-operated Head Start programs – is the customary level at which preschool curricular adoptions are made. The intervention we propose to test includes the curriculum and an implementation support package designed to make it possible for preschools to adopt, implement, and maintain the curriculum without requiring economically, culturally, or otherwise unfeasible accommodations by local programs. Furthermore, we want to test whether the implementation conducted under realistic circumstances on a broad scale produces the expected gains for children in those classrooms. The hypotheses and research questions to be addressed, therefore, are as follows:

Hypothesis 1. The math intervention will lead to increased mathematical knowledge in participating children by the end of the pre-kindergarten year.

Hypothesis 2. The training provided to prekindergarten teachers will alter their mathematics practices, in particular, to adopt a set of best mathematics practices.

Hypothesis 3. Causal influences of the intervention will be mediated by these best mathematics practices.

Research Question 1. Will intervention effects be similar across variations in geographical sites, program types, and child/family characteristics?

Research Question 2. Will the math intervention influence aspects of children's language and literacy skills?

Research Question 3. Will the hypothesized effects of the intervention on children's mathematical knowledge persist in elementary school?

Method

Participants. To be as representative of scale-up conditions as feasible, the preschool sample was selected to provide diversity in location, type of preschool program, teacher characteristics, and background and characteristics of the children and their families. The project was conducted in publicly funded public preschool programs serving urban, ethnically diverse families in California (CA) and predominantly rural Caucasian families in Kentucky (KY) and Indiana (IN). Both Head Start and state-funded prekindergarten programs in CA and KY/IN were included. These two types of programs provide the vast majority of preschool educational opportunities to low-income families in the US, but they differ in their administrative structures and income requirements for the populations they serve.

Teacher participants. Prekindergarten teacher participants included 48 treatment teachers (24 CA and 24 KY/IN) and 46 control teachers (24 CA and 22 KY/IN). The amount of teaching experience of treatment and control teachers was 16 years and 14 years, respectively in CA and 13 years and 13 years, respectively, in KY/IN. The percentage of treatment and control teachers with a BA degree or more was 62% and 48%, respectively, in CA and 52% and 64%, respectively, in KY/IN.

Child sample. Children were randomly selected in each classroom (mean, 7.9 per classroom) from the total number of consented children, balanced for age and gender, to participate in the experiment. All children were in the age range eligible for entrance into public kindergarten the following year. The pretest sample consisted of 744 normally developing prekindergarten children (367 in CA and 377 in KY/IN) from low-income families. Their mean age was 4.44 years (range, 3.86-4.99 years) at pretest at the beginning of their prekindergarten year. Mean age was 4.43 years for the 387 treatment children and 4.44 years for the 357 control children. Ethnic composition of the sample was 17.3% African American, 52.3% Caucasian, 18.0% Hispanic/Latino, 3.0% Asian American, and 9.4% multi-ethnic/other (see Figure 3); 91% of children were fluent in English and 9% primarily spoke in Spanish or a combination of Spanish and English (see Table 1). Attrition during the prekindergarten year was 10.1% overall (10.85%, treatment; 9.24%, control), and occurred primarily due to family mobility.

Design. The curriculum effects of interest are increases in mathematics achievement for prekindergarten children that are sustained as they progress through school. The smallest educational unit for which a prekindergarten curriculum could be feasibly implemented, however, is a classroom, and actual implementation at scale is likely to involve larger units, e.g., school districts or Head Start grantee agencies. The pre-kindergarten mathematics intervention was conducted in both Head Start and state-funded preschool programs within the greater Sacramento, California metropolitan area and in rural counties of Indiana and Kentucky outside the greater Louisville, KY metropolitan area[1]. Directors of all participating programs agreed to include all classrooms in random assignment. Thus, participation by head teachers was mandatory. Classrooms primarily serving a linguistic community other than English-or Spanish-speaking families were excluded. A total of 94 preschool classrooms (28 Head Start and 20 state-funded classrooms in CA and 26 Head Start and 20 state-funded classrooms in KY/IN) were randomly assigned in the school year prior to implementation using block randomization. Blocks were formed at the program level, with classrooms from Head Start and state-funded programs balanced by curriculum condition when possible in each state. Classrooms were randomly assigned to either the treatment condition (*Pre-K Mathematics* curriculum) or the business-as-usual control condition (existing classroom curriculum), with the constraint that classrooms at the same site could not be assigned to different conditions.

Protections against treatment diffusion. Treatment and control classrooms were located on different program campuses (Head Start centers or school sites) to minimize the potential for treatment diffusion in control classrooms. Furthermore, programs were asked to ensure that teachers in treatment and control classrooms did not substitute in or transfer to classrooms assigned to a condition different than their original one. Finally, periodic observations were made in all project classrooms using the Early Mathematics Classroom Observation (EMCO) and Generalized Fidelity instruments. This set of instruments has sufficient sensitivity to detect the use of *Pre-K Mathematics* activities or activities from another curriculum in a classroom. No evidence of treatment diffusion was found in control classrooms.

State	Program Type	Condition	Sites	Classrooms	Children
CA	Head Start	Treatment	9	14	99
CA	Head Start	Control	9	14	102
CA	State-Funded	Treatment	9	10	82
CA	State-Funded	Control	9	10	84
KY/IN	Head Start	Treatment	6	13	109
KY/IN	Head Start	Control	5	13	94
KY/IN	State-Funded	Treatment	8	11	97
KY/IN	State-Funded	Control	8	9	77
Total			63	94	744

Characteristics of classrooms at baseline. Baseline observations were made in treatment and control classrooms once during the year prior to implementation of the intervention and prior to informing teachers about their status as treatment or control teachers. Observers were blind to the condition to which a classroom had been randomly assigned. The general classroom curricula used by program participants included *Creative Curriculum* and locally developed "American eclectic" curricula. Administration of the Generalized Fidelity instrument revealed that no branded math curriculum was being systematically implemented in any classrooms at baseline. Control teachers were asked not to alter their usual classroom practices or curricular choices during the year of the intervention. They were told they could receive the same math training as treatment teachers after control children had finished preschool.

The Early Childhood Environment Rating Scale-Revised (ECERS-R) was used to measure the general quality of participating classrooms during the baseline year prior to implementation of the intervention. The overall classroom quality rating on this instrument can range from 1 (inadequate) to 7 (excellent). Overall scores of classrooms in CA and KY/IN fell slightly below a rating of good (5.0). Scores of treatment and control classrooms were 4.9 and 4.7, respectively, and were found not to be significantly different. This overall rating of project classrooms was similar to the rating of 4.9 found in the Head Start Family and Child Experiences Survey, a large national study of Head Start programs (Resnick & Zill, 1999).

Teachers' math practices were measured at baseline by administering the Early Mathematics Classroom Observation (EMCO). The mean number of minutes of math (MOM) support that project teachers provided per child per day was recorded. Teachers in treatment and control classrooms provided pre-kindergarten children with 11.3 minutes and 11.1 minutes of math support per day (total MOM), respectively, and did not significantly differ. Thus, random assignment resulted in sets of treatment and control classrooms of similar overall quality and amount of math support.

Intervention. A pre-kindergarten mathematics intervention was implemented in public preschool programs in California, Kentucky, and Indiana. The purpose of the intervention was to enrich mathematically children's preschool and home learning environments. Control teachers were instructed not to alter their classroom practices or curricular choices from business as usual. Treatment teachers added the math intervention to their current set of classroom curricula.

Curricular components. The intervention included the classroom and home components of *Pre-K Mathematics*. The classroom and home activities comprising these components were implemented according to a standard curriculum plan. Teachers also materially enriched the classroom by installing or enhancing an existing math learning center in their classroom and by providing children with opportunities to use supplemental math software (*DLM Express*). Children were allowed to choose to utilize the math center during free play periods and were directed to use the math software at least once per week.

The *Pre-K Mathematics* curriculum consisted of small-group mathematics activities with concrete manipulatives for use by teachers and children in classrooms and home mathematics activities and materials for use by parents and their preschool-age children in home settings. A teacher's manual provided a curriculum plan that linked small-group classroom activities to home activities and provided an implementation schedule. Curriculum units, consisting of sets of mathematically related activities, were (1) Counting and Number, (2) Understanding Arithmetic Operations (Fall Activities), (3) Spatial Sense and Geometry, (4) Patterns, (5) Understanding Arithmetic Operations (Spring Activities), and (6) Measurement and Data. Teachers were directed to present 26 small-group mathematics activities in accordance with the curriculum plan at a rate of approximately one new math activity per week. The week's math activity was conducted with groups of 4-6 children (typically groups of 4) twice during the week for approximately 15-20 minutes per group. Thus, the scheduled curriculum dosage per child was 52 small-group sessions (26 activities X 2 sessions in which the child participated). When teacher absences lasted longer than one week, an effort was made to have an internal facilitator or other program staff members conduct small-group activities. Child absences were handled, in so far as possible, by using Fridays as makeup days for missed small group activities.

Small-group activities were designed to be sensitive to the developmental needs of individual children. The curriculum included specific suggestions about how to scaffold children who experienced difficulty with part of an activity. Furthermore, downward (less-challenging) extensions were provided for children who were not ready for the initial (easiest) part of the activity, and upward (more-challenging) extensions were suggested for children who completed the activity with relative ease. Assessment Record Sheets, customized for each small-group activity, were an implementation tool provided for teachers to use to keep a record of individual children's problem solving during each activity. The curriculum plan provided review time (one review day each week and periodic review weeks) for teachers to review small-group math activities with children who had been absent or were continuing to experience difficulty with a particular activity.

Teachers sent 16 *Pre-K Mathematics* home activities to parents for use at home over the course of the year. Approximately once every 1-2 weeks teachers sent an activity kit home with children. This kit contained a brief narrative description of the activity, its purpose, materials required to conduct the activity, and a picture strip depicting how to conduct the activity and suggested questions to ask. A bilingual (English and Spanish) version of home math activities was provided to Spanish-speaking families. The mathematical content of home activities was similar to the content teachers were covering in classroom activities during the same time period.

Professional development of teachers. A trainer-of-trainers model of professional development was used to enable staff in preschool programs to implement *Pre-K Mathematics* under conditions of routine educational practice. Program trainers, hereafter referred to as internal facilitators, attended a *Pre-K Mathematics Trainers Institute* that project trainers (external facilitators) conducted. Internal facilitators were either members of a participating school district or Head Start program's permanent training staff or independent contractors, depending on the way programs routinely provided professional development for their teaching staff. Training was provided in two phases. Phase 1 focused on helping internal facilitators learn the *Pre-K Mathematics* curriculum and the implementation tools (curriculum plan, Assessment Record Sheet, and Math Mastery Form, see Appendix B). Phase 2 of the institute focused on techniques for providing on-site support and monitoring implementation fidelity and curriculum dosage in individual classrooms. Following completion of the institute, internal facilitators (1) provided on-site training or facilitation to treatment teachers every other week, including fidelity checks with feedback - a formative evaluation, (2) monitored teachers' implementation by administering the Fidelity Visit Checklist (see Appendix B), reviewing Assessment Record Sheets and Math Mastery Forms, computer activity records, and home activity distribution records, (3) helped teachers address classroom management issues or other problems that were interfering with implementation, (4) kept records of specific, dated facilitation activities using the Fidelity Visitation Record, (5) monitored child attrition in treatment classrooms, and (6) discussed arrangement of teachers' classrooms in order to optimize the location of the small group table(s), computers, and math center. The project schedule called for these support and monitoring visits to be conducted in treatment teachers' classrooms approximately every other week. Implementation records indicated that facilitators actually conducted, on average, 13.1 visits (range, 7-20 visits) per classroom during the school year. During these classroom visits, facilitators observed teachers' delivery of *Pre-K Mathematics* small group activities and provided feedback and training as needed. Also, facilitators examined children's progress on the Math Mastery form, observed children during use of the computer-based mathematics activities, and confirmed that *Pre-K Mathematics* home activities were being sent home to families on schedule.

External facilitators conducted co-fidelity visits with all internal facilitators to monitor the quality of on-site training provided by internal facilitators (see Fidelity Visit Checklist, Appendix B). These visits included reliability checks of small-group math fidelity evaluation, teacher record keeping, and other aspects of implementation that internal facilitators had been trained to support. Percent agreement between external and internal facilitators for overall fidelity scores was 81%.

Teachers in the treatment group were trained through a series of mandatory *Pre-K Mathematics* curriculum workshops by external and internal facilitators and through on-site support by internal facilitators. A 2-day introductory workshop was provided for teachers during late spring or early summer of the baseline year. The primary objectives of this workshop were to teach teachers 3 small-group math activities and to help them make plans for preparing their classroom for implementation, including selecting locations, furnishings, and materials for small-group math activities, a math center, and computer math activities. After the workshop, a facilitator visited the classroom to help plan how to prepare it for implementation, and teachers practiced the small-group math activities with older, kindergarten-bound children who would be leaving the program over the summer. Multiday fall (4 days) and winter (3 days) workshops were provided for teachers. Pre-K Mathematics instructional videos were used to demonstrate each small-group activity. Each demonstration was followed by hands-on practice of the activity in small groups of teachers with an internal facilitator who provided feedback and support. The fall workshop focused on Units 1-3 of the curriculum and the spring workshop focused on the remaining units. Also, assistant teachers were offered the opportunity to participate in a 1-day workshop to learn to implement the mathematics software used in the intervention. Approximately half took advantage of this opportunity.

Curriculum implementation measures. Several measures were used to document the quality and amount of curriculum implementation performed by teachers. A fidelity instrument, the Fidelity of Implementation Record Sheet was used by facilitators to measure the quality of treatment teachers' implementation of *Pre-K Mathematics*. Fidelity was evaluated on five dimensions of curricular implementation deemed essential by the curriculum developers: (1) adherence to the schedule of activities set in the curriculum plan, (2) preparation of all materials needed to conduct the activity, (3) delivery of the principal parts of small-group mathematics activities, (4) provision of developmental adjustments (scaffolding, downward extensions, and upward extensions) needed by individual children in the group, (5) making written assessments of individual children during the course of activities. Fidelity observations were made during facilitators' biweekly classroom visits throughout the year. During these visits, a trainer observed as a teacher conducted a small-group math activity and provided feedback afterward regarding any departures from fidelity. Fidelity ratings were scored independently by two raters (inter-rater agreement = 98%) and were based on the first small-group activity session observed during a fidelity visit – before formative feedback was given to the teacher. An overall score of .90 or greater (possible range, 0-1) was judged by the curriculum developers to be high. Overall fidelity of implementation across all classrooms was .91 (see Table 2). Thus, the curriculum was generally implemented as intended by the developers.

The Generalized Fidelity Instrument (Appendix B) was administered to document whether control teachers were implementing an organized mathematics curriculum, and if so, whether their implementation corresponded to dimensions of fidelity measured when treatment teachers implemented *Pre-K Mathematics*. It was found that no control teachers were implementing an organized mathematics curriculum.

The amount of exposure to the curriculum that children received – curriculum dosage – was obtained by tallying data from teachers' Assessment Record Sheets (ARS) and parents' Home Activity Checklists (Appendix B). Teachers recorded individual children's participation and performance in small-group math activities. The implementation goal was for each child to be exposed to each small group activity at least twice: 26 activities X 2 exposures = 52 planned doses per year. Children who experienced difficulties with particular activities could receive additional exposures as time permitted. The average number of doses children actually received was 44.2 (85% of the recommended number), excluding additional doses above two for the same activity. This average was below the maximum possible primarily due to child absences for which teachers did not subsequently compensate during review periods.

Information on parents' use of Pre-K Mathematics home activities was obtained through a home math activity checklist completed by parents. Parents were asked to report whether they had received these activities from their child's teacher and, if so, how often they used the activities with their child. On average, parents reported using 9.8 (61%) of the 16 Pre-K Mathematics home activities, with a range of 0-16 activities per family.

Teachers were encouraged to set a goal of bringing as many children to mastery on individual activities as time permitted. Teachers transferred ARS performance data onto a Math Mastery sheet that made it possible to monitor individual children's progress through the curriculum. Children were credited with having mastered an activity if they had received at least two doses of an activity and on the final exposure to the activity needed no help from their teacher. Teachers continued reviewing activities with children who had not mastered them. According to teachers' Math Mastery sheets, children mastered approximately 71% (18.4 of 26) of the small-group math activities; 18% of children mastered 0-12 activities, 32% mastered 13-18 activities, 32% mastered 19-24 activities, and 19% mastered 25-26 activities (also see Figure 4).

Children's mastery of activities predicted their posttest math scores. Controlling for pretest scores, regression analyses revealed that mastery of *Pre-K Mathematics* small-group activities significantly predicted posttest scores on both the CMA, $F\ (2, 342) = 161.24, p<.0001$, and the TEMA, $F\ (2, 342) = 294.04, p < .0001$.

Assessment procedures. The assessment battery included a combination of experimenter-developed and standardized child assessment instruments, and classroom and home learning environment measures.

Child assessment. In order to test the hypothesis that treatment children will develop more extensive mathematical knowledge than control children, child assessment instruments should be sufficiently broad to include diverse knowledge structures that comprise early mathematical cognition and are supported by the pre-kindergarten math curriculum. Two instruments, the Test of Early Mathematics Ability-3 (TEMA) and the Child Math Assessment (CMA) were used concurrently. The TEMA is a standardized measure of children's mathematical knowledge. Its content is restricted to informal and formal numerical abilities, including arithmetic problem solving (see Ginsburg & Baroody, 2003, for a detailed description). The CMA is a researcher-developed measure with better content validity than the TEMA. The CMA assesses informal mathematical knowledge across a broad range of concepts and skills that are developing during early childhood (see Psychometric Study above). Both instruments used tasks and materials that differed from those used in the intervention, and, therefore, the child math measures were not over-aligned with the math activities included in the intervention.

Three language and early literacy subtests from the Woodcock-Johnson III Tests of Achievement (WJ-III) were also included in the child assessment battery. The Letter-Word Identification subtest assessed word-decoding skills; preschool items required identification and pronunciation of written letters and words. The Spelling subtest assessed children's ability to write orally presented letters and words. A WJ-III language subtest, Understanding Directions, assessed children's ability to follow oral directions to point to different parts of pictures. The rationale for including it arose during debriefing of a group of treatment teachers who had implemented our mathematics intervention for two years in a prior project (e.g., Klein et al., 2008). When teachers were asked about any behavioral changes they had seen in treatment children while participating in small-group math activities, several teachers suggested or agreed that children became better able to follow verbal instructions.

The TEMA, CMA, and WJ-III subtests were administered to children individually in three separate sessions in fall and spring of the pre-kindergarten year. The pretest assessment was conducted prior to the time teachers began implementing the curricular intervention in treatment classrooms. The posttest assessment was conducted after the curricular intervention had been completed in treatment classrooms. The TEMA and CMA were administered individually to children in separate sessions (A and B). Session order for the math assessment was counterbalanced across children, with half receiving session A (TEMA) first and half receiving session B (CMA) first. Session C (WJ-III subtests) was presented last. Testing took place at children's preschools. Individual children were escorted from their classroom to a quiet room for the testing sessions. Children were assessed in math again in grades K and 1. Their elementary school teachers were not informed about the condition to which children had been assigned.

Classroom and home learning environment measures. A classroom observation was conducted in the baseline year to establish whether measured features of treatment and control classrooms were similar or whether an unhappy randomization had occurred by chance. Observers were blind to the condition to which classrooms were assigned. The baseline classroom observation included three instruments, the Early Childhood Environment Rating Scale-Revised (ECERS-R), the Early Mathematics Classroom Observation (EMCO), and the Numeracy Environment Checklist (NEC). The ECERS-R has been used in many public and private preschool programs to measure general program quality (e.g., National Association for the Education of Young Children, 2010). It was used in this study to establish whether the measured features of treatment and control classrooms were both similar to one another and, on average, similar to those of public preschool classrooms nationally. Finally, a teacher questionnaire was administered at baseline to collect demographic data on project teachers, including education and teaching experience, self-reported instructional emphases, and teacher beliefs about school readiness.

A classroom observation instrument, the Early Mathematics Classroom Observation (EMCO), was used to assess the nature and amount of mathematics support teachers' were providing in their classrooms. This instrument served two purposes. First, it was administered in the baseline year, prior to the introductory workshop, to establish whether random assignment of classrooms to condition had, in fact, resulted in treatment and control classrooms with similar math support at baseline. Second, the EMCO was administered in fall and again in spring of the scale-up intervention year to measure differences in math practices of treatment and control teachers. This made it possible to detect effects of the training provided to teachers on their math practices. The observation of math practices began during the arrival of children and ended during departure (half-day classrooms) or lunch (full-day classrooms). An observation was made of all teacher-participant activities in which there was mathematical content. The observer recorded the mathematical content, number of children present, and duration of the activity. Activities were categorized as (a) *focal* math activities in which the primary goal of the activity was children's mathematics learning (e.g., an activity focused on the names of numerals) or (b) embedded math activities in which the primary goal was non-mathematical (e.g., a cooking project in which children measured ingredients at one point). By recording the duration and number of children present during each teacher-participant math activity, the EMCO makes it possible to calculate the average amount of time children receive math support during a school day.

The Numeracy Environment Checklist (NEC) was also administered during classroom observations to catalog math materials in the classroom. Observers recorded whether classrooms had a math learning center and the number and types of child-accessible math materials (e.g., sets of blocks, board games with dice, number or shape books, math software, measurement tools, etc.) found in the math center or elsewhere in indoor or outdoor areas used by the class. The EMCO and NEC were administered twice (fall and spring) in the implementation year in treatment and control classrooms.

A home math environment questionnaire was administered during winter of the pre-kindergarten year to treatment and control families in California. It was administered only to intervention families in Kentucky/Indiana due to experimenter error. The principal purpose of the questionnaire was to provide data on the nature and amount of support for mathematical development children were receiving at home. The questionnaire asked parents to report how often their child had engaged in specific activities during the past week and whether the activity was done by the child alone, with other children, or with an adult. Activities were listed by type (e.g., books and language activities) and specific activities (story book; counting or shape book, etc.). Types of activities included books and language activities, store-bought games, activities that use technology, other educationally or mentally stimulating activities or toys, activities during the home routine, and activities away from home (excluding preschool).

Results

Child math outcomes in the pre-kindergarten year. Individual problems comprising CMA tasks were scored for accuracy (possible range, 0-1). Since the numbers of problems varied across tasks, a mean proportion correct on individual tasks was computed. Then, a composite mathematics score for each child was obtained by averaging across the mean scores for all tasks. CMA composite scores provided a mathematically broad measure of children's early mathematical knowledge at pretest and posttest (see Table 3). These scores were used in impact analyses described below. Children's raw scores on the TEMA provided a standardized measure of early mathematical knowledge at pretest and posttest (see Table 3). These scores were also used in impact analyses described below.

A set of analyses tested the main hypothesis (Hypothesis 1 above) that the math intervention would lead to increased mathematical knowledge in treatment children by the end of the pre-kindergarten year. These analyses required the use of multilevel models, because observations (pretest, posttest) were nested within students, students were nested within classrooms, and classrooms were nested within sites. SAS Proc MIXED software (SAS, 2010) was used to carry out all analyses of the effect of the intervention over time. Both a compound symmetric and a compound symmetric with heterogeneous variances for the repeated measures portion were tested. The compound symmetric estimates were used if the fit was not significantly worse for that model. Otherwise, the model allowing the two assessments to have different variances was used.

Effects of potential covariates and moderators (child gender and age at pretest, state, and type of preschool program) were tested to see if they accounted for significant variation in the model and were included in tests of the condition effect if they were significant. Then, the condition effect, its interaction with time, and any other significant interactions were added to determine the effect of the intervention and whether its effect was moderated by these covariates. Residuals for the final model were tested for deviations from normality to ensure that the mixed model was appropriate. There were no appreciable differences from normality of the residuals.

By centering the time variable at the posttest, level difference between the intervention groups could be determined at the posttest; the intervention by time interaction represented the difference between pretest and posttest by condition. A model centering at the pretest in order to determine that there were no differences between groups at the pretest was also run. Effects sizes were estimated by determining the difference between the two intervention groups divided by the pretest standard deviation.

CMA scores. A 3-level ANOVA of children's mean composite scores on the CMA at pretest and posttest revealed a significant Condition X Time interaction, $F(1, 666)=171.28$, $p<.0001$, (ES=.83). Treatment children's scores increased from 0.29 at pretest to 0.61 at posttest, an increase of 102%; control children's scores increased from 0.32 at pretest to 0.48 at posttest, an increase of 50% (see Figure 5). Thus, treatment children's gains were approximately double those of control children.

TEMA scores. A 3-level ANOVA of children's raw scores on the TEMA at pretest and posttest revealed a significant Condition X Time interaction, $F(1,663)=47.66$, $p< .0001$ (ES=.45). Intervention children's scores increased from 7.0 at pretest to 14.8 at posttest, an increase of 111%; control children's scores increased from 7.4 at pretest to 12.5 at posttest, an increase of 69% (see Figure 6).

Moderation of effects. Research Question 1 asked whether intervention effects were similar across child/family characteristics and variations in geographical sites (California or Kentucky/Indiana) and types of programs (state-funded pre-kindergarten and Head Start programs). A repeated measures analysis was used with measures of children at two time points (pretest and posttest of the pre-kindergarten year) and with children nested within teachers. Since variability in the measures was more at posttest than at pretest, an analysis specifying an unstructured variance-covariance matrix was used. Condition was used as the independent variable and either ethnicity or gender was added as a covariate and potential moderator by including the interaction of condition by ethnicity or condition by gender as well as interactions by time. For example, the ethnicity variable addressed whether the level of the outcome varied as a function of ethnicity while the ethnicity by condition interaction determined whether the effect of condition on the level of outcome varied as a function of ethnicity. Ethnicity by time addressed the relation of ethnicity to the change from pretest to posttest, while the condition by ethnicity by time interaction addressed whether the condition effect on the change over time was moderated by ethnicity.

The analysis revealed that neither gender nor ethnicity appeared to moderate the effect of condition during the pre-kindergarten year. On the CMA, a significant main effect for gender was found, with higher scores by girls (.4327) than boys (.4072), F (1, 666)=5.45, p< .02. Gender did not interact with condition. On the TEMA, a significant interaction of condition, time, and gender was found, F (1, 1308)=5.51, p< .02. Girls' gains (7.94 points) were marginally greater than boys' (7.21 points) in the treatment condition, but boys' gains (5.80 points) were marginally greater than girls' (4.97 points) in the control condition (see Figure 7).

There was no significant main effect of ethnicity or interaction involving ethnicity and condition on the CMA or TEMA. Ethnicity significantly interacted with time, with more change on the CMA observed in Caucasian children (fall, .3180; spring, .5517; change, .2337) than African-American children (fall, .2885; spring, .4958; change, .2073), t (1306)=2.35, p< .02 (see Figure 8).

Similarly, there were no significant effects of state or type of program and no significant interaction of time with state or type of program. This indicates that the intervention was robust across varied preschool contexts - Head Start and state preschool programs, or programs in California serving urban, ethnically diverse families and programs in Kentucky and Indiana serving predominantly rural white families.

Child language and literacy outcomes. Research Question 2 asked whether the math intervention would impact aspects of children's language and literacy abilities. Analyses of WJ-III W Scores for the Understanding Directions, Letter-Word ID, and Spelling subtests revealed no significant main effects of condition or interactions involving condition (see Table 4).

Teachers' mathematics practices. A set of analyses tested the hypothesis (Hypothesis 2 above) that training provided to prekindergarten teachers would lead teachers to alter their mathematics practices. Treatment teachers, but not control teachers, received specialized training for implementing the pre-k mathematics intervention. It was expected that this training would lead treatment teachers to adopt a set of best mathematics practices.

Intermediate effects on teachers' mathematics practices. EMCO data on teachers' mathematics practices in the baseline and implementation years were scored for characteristics of math activities and amount of time teachers spent engaged in math activities with children during a one-day period. The broadest temporal measure obtained was minutes of math per child across all types of math activities (total MOM). It was hypothesized that intervention teachers, relative to control teachers, would spend more time supporting children's mathematical development. During the baseline year treatment and control teachers spent similar amounts of time supporting math, 11.1 min and 11.3 min (total MOM), respectively. During the year of implementation of the intervention, however, treatment teachers spent more time supporting math than control teachers did, 25.8 min and 12.2 min (total MOM), respectively, $F(1,31)=30.21$, $p < .0001$.

The set of teacher-participant math activities (total MOM) was then differentiated into (a) activities conducted with small groups of children (small-group MOM); (b) activities conducted simultaneously with all children in the classroom (whole-group MOM); (c) activities with an explicit math focus, regardless of whether they were small- or whole-group activities (focal MOM); (d) activities without a math focus but that contained math as an embedded subcomponent, such as a cooking project that included some measurement (embedded MOM); (e) activities in which teachers provided scaffolding for children (scaffolded MOM); and (f) activities in which teachers did not provide scaffolding (non-scaffolded MOM). Note that some categories overlap.

It was hypothesized that intervention teachers, relative to control teachers would exhibit a set of best practices emphasized in the written *Pre-K Mathematics* curriculum and in the training provided to them (see parameter *a*, Figure 10). This set of best practices included greater use of small-group activities, focal activities, and scaffolding (see a, c, and e, above). Teachers' use of whole group activities, embedded activities, and non-scaffolded activities, which were not included in the curriculum or training, were also examined, but no predictions were made about the impact of the intervention on them. Analyses were conducted to compare the amount of time intervention and control teachers used these specific practices during the implementation year. As hypothesized, math practices of intervention teachers, relative to control teachers, included more small-group MOM, $F(1,31)=68.92$, $p < .0001$, more focal MOM, $F(1,31)=36.16$, $p < .0001$, and more scaffolded MOM, $F(1,31)=36.41$, $p < .0001$. In contrast, intervention and control teachers did not differ in whole-group MOM, embedded MOM, or non-scaffolded MOM, all *p*s > .05 (see Figure 9).

Classroom observations using the NEC indicated that 71% of treatment classrooms had dedicated math learning centers and 100% had computers, whereas and 39% of control classrooms had math learning centers and 89% had computers. Teacher records of use of math software indicated that teachers in 89% of treatment classrooms had all pre-kindergarten children use math software on a regular basis.

Parental support for math at home. The home math environment questionnaire provided a one-week snapshot of treatment children's mathematical learning environment at home (see Table 11). Parents reported conducting approximately 10 focal math activities (indicated by asterisks in the table) per week and at least 26 explicit language and literacy activities. (Parents did not state the specific content for some activities, such as watching an educational TV program, which likely contained language and literacy content.)

Mediation of causal influences. Mediation analyses attempt to determine the degree to which an active ingredient or intervention condition (causal influence) is transmitted from the cause to the effect by an intervening variable, the mediator. In the present situation, the active ingredient is the mathematics content of *Pre-K Mathematics* activities (see *Intervention Condition* in Figure 10), the effect is the child's posttest score on a math assessment (see *Child Math Outcomes* in Figure 10), and the proposed mediators are measures of teacher's use of strategies thought to promote mathematical learning and development (see *Best Math Practices* in Figure 10).

It was predicted that causal influences would be mediated by changes in teachers' mathematics practices (Hypothesis 3 above). According to Baron and Kenny (1986), evidence for mediation requires several steps. First, the causal influence (intervention condition) should be related to the effect (child math outcomes) (see parameter *c*, Figure 10). This requirement was met by obtaining support for Hypothesis 1. Second, the causal influence (intervention condition) should be related to the mediator (best math practices) (see parameter *a*, Figure 10). This requirement was met by obtaining support for Hypothesis 2. Third, the mediator (best math practices) should be related to the effect (child math outcomes) (see parameter *b*, Figure 10). To address this requirement, a set of analyses was conducted to determine whether specific mathematics practices by teachers predicted child outcomes in mathematics. The model used for the analysis included children's pretest score as a covariate. Total MOM, small-group MOM, focal MOM, and scaffolded MOM predicted child outcomes as measured by the CMA and TEMA (see Table 5), but whole-Group MOM, embedded MOM, and non-scaffolded MOM did not (see Table 6). Thus, duration of mathematics instruction, and in particular time devoted to the set of best mathematics practices, predicted child outcomes in mathematics.

A final condition is that the effect of the causal influence should be reduced or eliminated when the mediator is included in the model. Assessment of mediation, however, is complicated by the fact that the main way of assessing the mediation effect is by considering the product of two parameters, the effect of the intervention condition on best math practices (parameter *a*) and the effect of best math practices on child math outcomes (parameter *b*). Even if parameters *a* and *b* are normally distributed, the product may not be. Bootstrapping techniques give a way of estimating the meditational effect without the need to assume normality in the sampling distributions. Our analysis also includes parameter *c*, which represents the direct effect of intervention condition on child math outcomes (specifically, TEMA raw scores), while controlling for best math practices. Partial mediation is implied when the direct effect of the causal influence is still significant even after the mediator is included in the model, assuming the effect of the mediator is also significant. If the direct effect is insignificant when the mediator is included, then total mediation is implied.

For each mediator - each particular math practice - two models were run. The first modeled the effect of intervention condition on the math practice using a mixed model. The second modeled the effect of the intervention on child math outcomes but included the math practice as a covariate in the model. This second model also included the child math pretest

(TEMA raw score) as a covariate and controlled for type of program (Head Start or state-funded preschool) and state (CA or KY/IN) and interactions. Each model was run on 5000 replicate samples taken with replacement from the original data set. Then the 95% confidence interval for each parameter was determined from the 5000 replicate samples. Means and standard deviations for the estimation of effect size are given below. Due to coding of the intervention effect (control = 1, intervention = 0), parameters relating intervention condition to other effects are negative, whereas parameters relating the mediation effect are positive. Therefore, a 95% confidence interval for the indirect effect and parameters *a* and *c,* with both endpoints being negative, indicate a positive effect of the intervention.

Results were obtained for 5000 bootstrap samples using total MOM as the mediating factor (see Table 7). The effect of the intervention on total MOM had mean of -8.728 (SD = .151). The 95% confidence interval (CI) for parameter *a* in the model did not contain zero (-9.001, -8.366) indicating a significant effect of the intervention on total minutes of math. Likewise, the 5 % CI for parameter *b* did not contain zero (.016, .094), indicating that the mediator effect on the TEMA was significant when condition was controlled. The 95% CI for the indirect effect (*a***b*) also did not contain zero (-0.826, -0.144) indicating a significant result and providing evidence that the effect of the intervention was at least partially mediated by total minutes of math. Given that the effect of the intervention on the TEMA, while controlling for the mediator, was still significant (95% CI = -2.722, -0.630), the significant indirect effect does indicate partial mediation. Based on the standardized effect sizes, the indirect effect accounts for about 47% of the total intervention effect on the TEMA, (-.481/.174)/((-.481/.174)+(-1.680/0.530)).

Results were also obtained for 5000 bootstrap samples using small-group MOM as the mediating factor (see Table 8). The effect of the intervention on the mediator was significant as the 95% CI did not contain zero (-8.18, -7.70). Likewise the effect of the mediator on the outcome controlling for intervention condition was significant (.008, .150). The fact that the indirect effect 95% CI did not contain zero (-1.202, -.065) nor did the 95% CI for the effect of the intervention on the TEMA controlling for the mediator (-2.715, -.493), there was evidence of partial mediation of the intervention on the TEMA by small-group MOM. Examination of the effect sizes indicates that 43% of the total effect of the intervention on the TEMA was mediated by small-group MOM.

Results were similarly obtained for focal MOM (see Table 9). Both the effect of the intervention on the mediator (-8.000, -7.486) and the effect of the mediator on the TEMA (0.013, 0.114), controlling for the intervention, were significant. There was a significant indirect effect of the intervention on the TEMA through focal minutes of math (-0.884, -0.100), indicating evidence of mediation, but also a significant effect of the intervention on the TEMA when controlling for the mediator (-2.745, -0.638), indicating that the mediation was partial. Examination of the effect sizes indicated that about 44% of the total effect of the intervention on the TEMA was though focal minutes of math.

Finally, results were similarly obtained for scaffolded MOM (see Table 10). Both the effect of the intervention on the mediator (-7.945, -7.394) and the effect of the mediator on the TEMA (.022, .115), controlling for the intervention, were significant based on the 95% CI. In addition, the indirect effect of the intervention on the TEMA through scaffolded MOM (-2.681, -.527) and the direct effect of the intervention on the TEMA controlling for the mediator (-0.892, -0.167) were significant, indicating evidence of partial mediation. Examination of the standardized effect sizes indicated that almost 50% of the total effect of the intervention on the TEMA was through its impact on scaffolded MOM.

Kindergarten and first grade follow up. The next set of analyses examined whether the effects of the intervention on children's mathematical knowledge persist into elementary school (Research Question 3). Children's raw scores on the TEMA at prekindergarten pretest (time 1), prekindergarten posttest (time 2), kindergarten (time 3) and first grade (time 4) were analyzed using a repeated measures mixed model assuming compound symmetry (see Table 12). There were significant effects of type of program, $F(1, 1760) = 5.58$; $p < .02$, condition, $F(1, 1760) = 4.05$; $p < .05$, and time, $F(3, 1760) = 3285.9$; $p < .0001$, and a significant time by condition interaction, $F(3, 1760) = 9.64$; $p < .0001$.

Tests of the difference in scores between conditions were conducted at each time point. It was found that conditions did not differ significantly at time 1, but did differ by 2.36 points (SD=0.53) at time 2, $t(1760) = 11.07$; $p < .001$. A difference of 1.18 points (SD=0.26) at time 3 was of borderline significance, $p = .051$, and a difference of 0.68 points (SD= 0.15) at time 4 was not significant. In Figure 11, it can be seen that scores of the treatment and control groups diverged during prekindergarten and began to converge in grades K and 1.

Uniform Report Card ratings. Teachers were blind to the condition to which children had been assigned. ANOVAs of teacher ratings of children's overall mathematics performance in kindergarten and first grade revealed no significant main effect of condition or interaction with condition.

Math instruction in grades K and 1. The Math Instruction Survey was used to obtain data on elementary school mathematics instruction provided to project children. Kindergarten teachers reported teaching mathematics 32 min per day in whole groups and 13 min per day in small groups. First grade teachers reported teaching mathematics, 44 min per day in whole groups and 14 min per day in small groups. Regression analyses, controlling for prekindergarten posttest TEMA scores, found that math instructional time did not predict kindergarten children's TEMA scores. Thus, the condition differences in kindergarten TEMA scores did not appear to be the result of the amount of math instruction children received in kindergarten. Finally, regression analyses, controlling for kindergarten TEMA scores, found that math instructional time did not predict first grade children's TEMA scores.

Summary and Conclusions

Classroom observations, made during the baseline year that preceded the intervention, revealed that teachers were providing children with only 11 minutes of mathematics support per day. Furthermore, teachers were not implementing an organized mathematics curriculum of any kind. During the intervention year, a professional development model was employed that included seven days of workshops and biweekly on-site training by facilitators. This training was sufficient for treatment teachers to implement Pre-K Mathematics with a high degree of fidelity (91%), to deliver a good dosage of small-group math activities in their classrooms (85% of the recommended amount), and to get parents to provide a fair-to-good dosage of home math activities (61% of the recommended amount). Teachers, with support from internal facilitators, monitored and recorded children's progress in math across the pre-kindergarten year. They found that children mastered, on average, 71% of small-group math activities.

The principal hypothesis (Hypothesis 1) of the study was that the math intervention would lead to increased mathematical knowledge in participating children by the end of the pre-kindergarten year. This hypothesis was clearly supported both by CMA scores (ES = .83) and by TEMA scores (ES=.45). Treatment children's gains on the CMA over the pre-kindergarten year were approximately double those of control children. Furthermore, the effect of the *Pre-K Mathematics* intervention was found to persist to a significant degree in kindergarten but not first grade (Research Question 3).

The intervention was effective across variations in geographical sites (urban programs in California and rural programs in Kentucky and Indiana), program types (Head Start and state-funded Pre-K programs), and child/family ethnicity (Research Question 1). There was a significant main effect of gender, favoring girls, on the CMA, and gender interacted significantly with condition and time on the TEMA. Girls made greater gains than boys in the treatment condition; whereas, boys made greater gains than girls in the control condition. The math intervention did not appear to have an effect on aspects of literacy (letter-word identification or spelling) and language (understanding directions) in children (Research Question 2).

Thus, the math intervention had significant effects on children's math outcomes that persisted into elementary school and were robust across variations in geographical location, program type, and child/family ethnicity. There was some evidence that the intervention benefitted girls somewhat more than boys.

Another hypothesis of the study (Hypothesis 2) was that the training provided to treatment teachers would alter their mathematics practices. Specifically, teachers would learn to utilize a set of best mathematics practices consisting of small-group activities, focal activities, and scaffolding. As predicted, treatment teachers, relative to themselves at baseline and relative to control teachers, spent significantly more time, in general, supporting children's mathematical learning and development, and more time, in particular, utilizing the set of best mathematics practices that had been emphasized in their professional training. Use of these practices predicted children's posttest math scores.

Treatment teachers spent less time utilizing a set of (theoretically) non-optimal mathematics practices, including whole group activities, embedded activities, and non-scaffolded activities. Untrained control teachers primarily used these non-optimal practices to support math. Analyses revealed that, unlike the set of best practices, these non-optimal practices did not predict children's posttest math scores. Parental reports of math support they provided at home, other than *Pre-K Mathematics* home activities, revealed little support for math (10 activities per week) and far less than language and literacy support ($\geq$26 activities).

A final hypothesis (Hypothesis 3) of the study was that causal influences of the intervention would be mediated by changes in teachers' mathematics practices. Specifically, aspects of adult-guided mathematical activities - a set of best mathematics practices - would transmit causal influences by engaging children's mathematical cognition. Mediation analyses supported the hypothesis that these practices played a mediating role.

Next Steps. Following completion of the pre-kindergarten year of the Main Study, a Sustainability Study was commenced in the same treatment classrooms. The principal objective was to determine whether the Head Start and state-funded public preschool programs and their administrators, trainers, and teachers were able to sustain implementation of the *Pre-K* Mathematics intervention in an effective manner beyond the initial year in which they received professional development.

Sustainability Study

Hypothesis and Research Questions

The Main Study demonstrated that the Pre-K Mathematics intervention was effective when implemented at a customary level of scale for curriculum adoptions by public preschool programs. This study also demonstrated that the effects of a mathematics intervention in pre-kindergarten persist into kindergarten. An important subsequent question is whether public preschool programs are able to sustain their implementation with sufficient fidelity and dosage that continues to provide educational benefits to the children in their classrooms. To address this issue of sustainability, teachers randomly assigned to the treatment condition in the Main Study were expected to continue using the curriculum for an additional year. Although project staff monitored teachers' implementation of the curriculum, program staff members (teachers, internal facilitators, and program administrators) were responsible for ensuring continued use of the curriculum. Except where otherwise noted, the measures and data collection procedures mirrored those used in the Main Study, thus allowing for comparisons between the quality of implementation in Year 1 and Year 2, child outcomes in pre-kindergarten and kindergarten for the Sustainability treatment group and the Main Study control group, child outcomes in pre-k and kindergarten for the Main and Sustainability treatment groups, and treatment teachers' practices in Year 1 and Year 2. The hypothesis and research questions that were addressed included the following:

Hypothesis 1. The math intervention will lead to increased mathematical knowledge in participating children by the end of the pre-kindergarten year.

Research Question 1. Will child outcomes be similar in the Main Study and Sustainability Study at the end of the pre-kindergarten year?

Research Question 2. Will the hypothesized effects of the intervention on children's mathematical knowledge persist in elementary school?

Research Question 3. Will teachers sustain their implementation of the intervention with a comparable degree of fidelity and similar dosage levels as observed in the Main Study?

Research Question 4. Will teachers' mathematics practices in the classroom remain similar to those observed in the Main Study?

Research Question 5. Will the math intervention influence children's ability to follow verbal directions?

Method

Participants. Of the 48 treatment teachers from the Main Study, 39 (20 CA and 19 KY/IN) continued teaching at the pre-kindergarten level during the subsequent year in which the Sustainability Study was conducted. All 39 continued implementing the math intervention. Of the nine teachers who did not continue teaching in pre-kindergarten classrooms, three accepted positions teaching at other age or grade levels, two were promoted to supervisory positions, two took maternity leave, one become a high school counseling assistant, and one moved out of state. The mean teaching experience of the 39 remaining teachers was 17 years in CA, and 13 years in KY/IN. In CA, 65% of teachers had a B.A. degree, 20% had an A.A. degree, and 15% had less than an A.A. degree, and in KY/IN, 52% had a B.A. degree, 16% had an A.A. degree, and 32% had less than an A.A. degree.

Child sample. Children were randomly selected from the participating classrooms (mean, 8.4 per classroom) yielding a sample of 326 prekindergarten children (171 CA and 155 KY/IN), balanced for age and gender. The sample was selected using the same sampling procedures and eligibility criteria as described in the Main Study. Their mean age at the time of pretest in prekindergarten was 4.37 years (range, 3.82-5.04 years). The ethnic composition was 14.1% African American, 20.9% Hispanic/Latino, 3.0% Asian American, and 9.2% was multi-ethnic/other (see Figure 12). The majority, or 86.2% of children spoke English fluently, and the remaining 13.8% spoke either Spanish or a combination of English and Spanish (see Table 13). Attrition during the prekindergarten year was 11%, and occurred primarily due to family mobility.

Design. Classrooms remained in the same conditions as in the Main Study. That randomization resulted in the assignment of 48 classrooms to the treatment condition and 46 to the control condition. Teacher attrition resulted in the loss of one site and nine classrooms from the first year of implementation (Main Study) to the second year (Sustainability Study).

State	Program Type	Condition	Sites	Classrooms	Children
CA	Head Start	Treatment	8	11	95
CA	State-Funded	Treatment	9	9	76
KY/IN	Head Start	Treatment	6	9	72
KY/IN	State-Funded	Treatment	8	10	83
Total			31	39	326

Intervention. Teachers implemented the classroom and home components of *Pre-K Mathematics*, as described in the Main Study section for a second year. Project staff conducted a one-day training with internal facilitators at the beginning of the school year to review the instruments and procedures used for fidelity observations. Teachers participated in two days of refresher workshops to practice a sub-set of activities that were identified as more difficult to implement and to review procedures for documenting dosage and monitoring children's progress in mastering the mathematics activities. Project staff attended workshops, but internal facilitators had the responsibility of direct training of teachers on curriculum activities in small groups. Teachers were also given videos of *Pre-K Mathematics* small-group activities to use as a reference to help sustain the implementation with fidelity. Three programs chose to hold one day of the refresher training in the fall and one day in the spring, and one program elected to hold both days in the fall. All 39 teachers participated in the first day of the workshop, and 38 teachers participated on the second day. Internal facilitators conducted, on average, 9.9 fidelity visits per classroom (range 5-21) across the year. During these visits, teachers' implementation fidelity, dosage, and progress monitoring were checked, and on-site training and assistance were provided as needed.

Curriculum implementation measures. Implementation measures and data collection procedures were the same as those used in the Main Study. To determine whether teachers sustained their implementation of the intervention at a level comparable to their implementation in the Main Study (Research Question 3), comparisons of implementation fidelity and curriculum dosage were made. Teachers' overall fidelity of implementation across all classrooms remained high, .93, during the second year of implementation (see Table 14). This indicates that the quality of teachers' implementation continued to conform to what was intended by the developers (see Table 15 for a comparison of fidelity in years 1 and 2).

As in the Main Study, the implementation goal was for each child to be exposed twice to each of 26 curriculum activities for a total of 52 planned doses per year. However, teachers in one of the four preschool programs were not able to complete the final curriculum activity prior to posttest assessments due to school days lost to inclement weather. Thus, the 26^{th} activity was not included in analyses of implementation data. The average number of doses of small-group math activities that children actually received was 43.8 (88% of the 50 included in our analyses, and 84% of the recommended 52), excluding additional doses above two for the same activity. This indicates that the percentage of the recommended dosage that teachers actually delivered in the classroom was similar, and therefore sustained, during their second year of implementation (see Table 15).

Information on parents' use of *Pre-K Mathematics* home activities was obtained through a home math activity checklist completed by parents. Parents were asked to report whether they had received these activities from their child's teacher and, if so, how often they used the activities with their child. On average parents reported using 8.7 (54%) of the 16 Pre-K Mathematics home activities, with a range of 0-16 activities per family. This was similar, albeit somewhat lower, than the amount of usage reported by parents in the Main Study.

According to teachers' Math Mastery sheets, children mastered approximately 65% (16.2 of 25) of the small-group math activities (see Figure 13). Overall, the number of curriculum activities mastered was similar for the Main and Sustainability Study sample compared to the Main Study sample.

Assessment procedures. The measures used in the Sustainability Study were identical to those used in the Main Study, with the exception that the WJ-III Letter/Word Identification and Spelling subtests were eliminated. All assessment procedures followed in the Main Study were followed in the Sustainability Study. See the Main Study section above, for a description of all measures and procedures.

Child assessments in prekindergarten. Children's mathematical knowledge was assessed using the CMA and TEMA in the fall and spring of the prekindergarten year. As in the Main Study, the curricular intervention did not begin until all children had been pretested in the fall, and the intervention commenced just prior to posttest in the spring. The WJ-III Understanding Directions subtest was again administered at both time points.

Classroom and home learning environment measures. Teachers' mathematics practices were observed twice during the school year (fall and spring) using the EMCO. Classroom math materials were also catalogued using the NEC during these observations. Finally, information about the home learning environment was obtained by administering a home math environment questionnaire to parents in the winter of the pre-kindergarten year.

Kindergarten follow-up. To determine whether the effects of the intervention were sustained over time, the TEMA was administered in the spring of the kindergarten year. Kindergarten teachers were also asked to evaluate children's mathematical skills by completing a Uniform Report Card for each child. Teacher ratings on the Uniform Report Card yielded two types of scores: (1) teacher's rating of children's overall mathematical ability, and (2) the mean rating of children's ability on 14 skills across the domains of number, arithmetic, space and geometry, measurement, and pattern knowledge. Finally, teachers were asked to complete a Math Instructional Survey to provide information on the mathematical practices utilized in their classrooms.

Results

Child math outcomes in the pre-kindergarten year. As in the Main Study, CMA composite scores and TEMA raw scores were also used in impact analyses. A set of analyses tested the main hypothesis (Hypothesis 1 above) that the math intervention would lead to increased mathematical knowledge in treatment children by the end of the pre-kindergarten year. These analyses required the use of multilevel models, because observations (pretest, posttest) were nested within students, students were nested within classrooms, and classrooms were nested within sites. SAS Proc MIXED software (SAS, 2010) was used to carry out all analyses of the effect of the intervention over time. Both a compound symmetric and a compound symmetric with heterogeneous variances for the repeated measures portion were tested. The compound symmetric estimates were used if the fit was not significantly worse for that model. Otherwise, the model allowing the two assessments to have different variances was used.

Effects of potential covariates and moderators (child gender and age at pretest, state, and type of preschool program) were tested to see if they accounted for significant variation in the model and were included in tests of the condition effect if they were significant. Then, the condition effect, its interaction with time, and any other significant interactions were added to determine the effect of the intervention and whether its effect was moderated by these covariates. Residuals for the final model were tested for deviations from normality to ensure that the mixed model was appropriate. There were no appreciable differences from normality of the residuals.

By centering the time variable at the posttest, level difference between the intervention groups could be determined at the posttest; the intervention by time interaction represented the difference between pretest and posttest by condition. A model centering at the pretest in order to determine that there were no differences between groups at the pretest was also run. Effects sizes were estimated by determining the difference between the two intervention groups divided by the pretest standard deviation.

CMA scores. To evaluate the effectiveness of the intervention during the sustainability year, children's scores on the CMA and TEMA over the prekindergarten year were compared with those of control children from the Main Study. A 3-level ANOVA of children's CMA scores, with children nested within classrooms within sites, revealed no differences at pretest but a significant condition by time interaction, $F(1, 527) = 68.82$, $p < .0001$ (ES =.70), indicating a greater increase in mathematical knowledge over the year among the sustainability treatment children compared to control children (see Figure 14). Treatment children's scores increased from .30 at pretest to .58 at posttest, an increase of 93%, compared to the 50% increase for control children described in the Main Study (see Table 16). Differences between the mean scores of treatment and control children were greater for children in state preschool than for children in Head Start, $F(1, 527) = 5.32$, $p < .05$, with an effect size of .89 in state preschool compared to .50 in Head Start. There were no differences by program type in KY, but in CA, state preschool children outperformed children in Head Start, $F(1,522) = 34.57$, $p < .05$.

TEMA scores. A 3-level ANOVA of children's TEMA scores, with children nested within classrooms within sites, also revealed a significant condition by time interaction favoring treatment children, $F =(1, 527) = 34.57$, $p < .0001$ (ES = .45) (see Figure 15). There were no significant interactions including program type, but there was a condition by state interaction. Means and standard deviations at pretest and posttest are presented in (see Table 16). Differences between the mean scores of treatment and control children were greater for children in KY than for children in CA, $F(1, 527) = 3.93$, $p < .05$, effect size = .60 in KY and .30 in CA. Finally, there were no differences by condition at pretest, children's pretest scores were less related to posttest scores in CA than in KY, $F(1, 522) = 8.58$, $p < .01$.

Comparison of math outcomes during teachers' first and second year of implementation. The next set of analyses examined whether child math outcomes were similar in the Main and Sustainability Studies (Research Question 1). Mixed (multilevel) models, with students nested within classrooms within sites, were used to compare CMA and TEMA posttest scores between Main Study treatment children in teachers' first year of implementation and the Sustainability Study treatment children in teachers' second year of implementation. There were no significant differences between the two treatment groups on the CMA or TEMA at pretest or posttest (see Figures 16 and 17).

WJ-III Understanding Directions. The next set of analyses examined whether treatment children became better able to follow verbal directions as assessed by the Understanding Directions subtest of the WJ-III (Research Question 5). A 3-level repeated measures ANOVA, with children nested within classrooms within sites, revealed main effects for age, $F(1, 606) = 200.69$, $p < .0001$, with older children scoring higher than younger children, and state, $F(1, 606) = 10.31$, $p < .005$, with children in KY scoring higher than children in CA. There was also a strong condition by time interaction, $F(1, 606) = 20.27$, $p < .0001$ (ES = .51). Although children's mean scores at pretest did not differ by condition, children in the treatment group made greater gains over the course of the year compared to children in the control group (see Table 17).

Teachers' mathematics practices. A 3-level ANOVA revealed that treatment teachers continued to provided more overall support for math in the Sustainability Study year compared to control teachers, $F(1,77) = 19.394$, $p < .0001$, based on the total minutes of math (MOM) observed. Treatment teachers continued to use the set of best math practices (Research Question 4). Specifically, they provided more small-group MOM, $F(1,77) = 74.176$, $p < .0001$, focal MOM, $F(1,77) = 24.305$, $p < .0001$, and scaffolded MOM, $F(1,77) = 26.040$, $p < .0001$ compared to control teachers. Again, there were no significant differences by condition in the amount of whole-group MOM, embedded MOM, or non-scaffolded MOM, all *p*s > .05.

The overall amount of math support (total MOM) provided by treatment teachers did not differ significantly between the two years of implementation (see Figure 18). A series of 3-level ANOVAs also revealed no significant differences in the amount of small-group MOM, focal MOM, scaffolded MOM, whole-group MOM, embedded MOM, or non-scaffolded MOM from the Main to Sustainability years. This indicates that teachers' mathematics practices remained stable across both studies.

Classroom observations using the NEC found that 80% of treatment classrooms had dedicated math learning centers, compared to 74% of these same treatment classrooms in the Main Study. Computers were present in all classrooms, and math software was available for the children to use in 77% of these.

Home environment. The Home Learning Environment checklist included in the parent questionnaire was used to obtain a snapshot of the support for early mathematical development children are receiving in the home. Parents were asked how many times in the past week their child engaged in specific educational activities in the home. The means and standard deviations for each activity are presented in Table 18. As in the Main Study, there was no evidence that participation in these activities at home moderated the effects of the curriculum intervention.

Kindergarten follow up. The next set of analyses examined whether the effects of the intervention on children's mathematical knowledge persist into elementary school (Research Question 2). The impact of the intervention on children's mathematical knowledge one year after implementation was evaluated by comparing the TEMA scores of treatment and control children obtained in the spring of the kindergarten year (see Table 19). Children's raw scores on the TEMA at prekindergarten pretest (time 1), prekindergarten posttest (time 2), and kindergarten (time 3) were analyzed using a repeated measures mixed model. There were main effects for condition, $F(1,1398) = 21.93$, $p < .0001$, indicating that treatment children scored higher than control children across the three time points, and time, $F(1,1398) = 2727.86$, $p < .0001$, indicating that children's scores increased over time (see Figure 19). There was also a significant condition by time interaction, $F(1, 1398) = 34.65$, $p < .0001$. Tests between condition at each time point revealed no differences by condition at time 1, but treatment children demonstrated significantly greater mathematical knowledge at time 2, $F(1, 1398) = 20.37$, $p < .0001$, and time 3, $F(1,1398) = 38.44$, $p < .0001$.

Mixed models were used to compare the effects of the pre-kindergarten intervention at the end of kindergarten in the Main Study and Sustainability Study samples. Children in the Sustainability treatment group scored significantly higher than the Main Study treatment group on the TEMA, $F(2,377) = 10.60$, $p < .0001$, indicating that the effects of the treatment one year after commencement were stronger for children in the Sustainability treatment sample.

Uniform Report Card ratings. Teacher ratings of children's mathematical ability on the Uniform Report Card provide further evidence that the impact of the intervention, when implemented by prekindergarten teachers in their second year, was sustained through the kindergarten year. Ratings of treatment and control children's overall mathematical ability were compared using a three-level ANOVA. Kindergarten teachers were blind to children's condition assignment in pre-k and gave higher ratings of children's overall math ability to treatment children (M=2.18, SD=.55) than to control children (M=2.07, SD=.64), $F(1,545)=5.77$, $p < .05$. Teachers also rated children's ability on 14 mathematical skills, spanning the domains of number, arithmetic, space/geometry, measurement, and pattern knowledge. The mean rating across these skills was also higher for treatment children (M=2.00, SD=.46) compared to control children (M=1.82, SD=.50), $F(1,547)=21.245$, $p < .0001$.

Math instruction in kindergarten. The Math Instruction Survey was used to obtain data on mathematics instruction provided to project children. Kindergarten teachers of children in the Sustainability sample reported teaching mathematics, on average, 32 minutes per day (range, 0 min - 90 min) in whole groups, 16 minutes per day (range, 0 min - 48 min) in small groups, and 7 minutes per day (range, 0 min - 40 min) to individuals. The amount of math instruction provided by teachers in whole groups, small groups, or to individuals was not significantly correlated with children's TEMA scores in Grade 1 when children's TEMA scores in kindergarten were controlled for.

Summary and Conclusions

Sustainability of implementation. Treatment teachers were able to implement *Pre-K Mathematics* with a high degree of fidelity (93%), to deliver a good dosage of small-group math activities in their classrooms (88% of the recommended amount), and to get parents to provide a fair-to-good dosage of home math activities (54% of the recommended amount). Comparison of teachers' fidelity and dosage levels during their first year of implementation (Main Study) and their second year (Sustainability Study) revealed similar – and acceptable - levels both years (Research Question 3). Teachers, with support from internal facilitators, monitored and recorded children's progress in math across the pre-kindergarten year. They found that children mastered, on average, 65% of small-group math activities. As in the Main Study, treatment teachers, relative to control teachers, spent significantly more time supporting children's mathematical learning and, in particular, utilizing the set of best mathematics practices that had been emphasized in their professional training (Research Question 4).

Child math outcomes. The principal hypothesis (Hypothesis 1) of the study was that the math intervention would lead to increased mathematical knowledge in participating children by the end of the pre-kindergarten year. This hypothesis was clearly supported both by CMA scores (ES = .70) and by TEMA scores (ES=.45). As in the Main Study, treatment children's gains on the CMA over the pre-kindergarten year were almost double those of control children. Comparison of children's pre-kindergarten math outcomes during teachers' first year of implementation (Main Study) and their second year (Sustainability Study) revealed similar – and significant - outcomes both years (Research Question 1). Furthermore, the effect of the *Pre-K Mathematics* intervention again was found to persist to a significant degree in kindergarten (Research Question 2). Treatment children, relative to control children, had both higher TEMA scores and higher math grades on the Uniform Report Card in kindergarten.

Other child benefits. There was also evidence that children benefitted in another way from participation in the math intervention. Significant child outcomes on the Understanding Directions subtest of the WJ-III (ES = .51) indicated that treatment children showed more improvement than control children in following verbal directions from adults. This effect was found during teachers' second year of implementation but during their first year, which suggests that teachers may have become more effective over time in their ability to help children develop this important capacity.

Psychometric Study of the Child Math Assessment

Background and Need

Over the past decade, educators have become increasingly concerned about the low level of mathematics performance of U.S. students on national and international assessments. Moreover, there is a persistent gap in mathematics achievement at all grade levels between students from low-income and minority backgrounds and their more advantaged peers. Compelling evidence indicates that this socioeconomic-related gap in mathematics achievement emerges during early childhood and, if not addressed through effective early math interventions, this gap will become more pronounced over the elementary school years. These findings point to the need to provide preschool children, especially at-risk children from low-income backgrounds, with effective math interventions to support their development of early mathematical knowledge. However, evaluating the impact of early math interventions requires reliable and valid assessment instruments to measure growth in young children's mathematical knowledge.

A survey of available instruments revealed few valid measures that are appropriate for preschool children, and those that have been used to measure early mathematical competencies such as the Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001) and the Test of Early Mathematics Ability (Ginsburg & Baroody, 2003) have significant limitations. The Applied Problems subtest of the Woodcock-Johnson III has been used as an early math measure in a number of national studies, but it does not appear to be based on current research on mathematical development. It contains very few items that measure the informal mathematical knowledge of young children, it jumps abruptly to items that require formal (i.e., symbolic) mathematical thinking, and it is subject to ceiling effects with preschool children. The Test of Early Mathematics Ability (TEMA-3), on the other hand, contains many items that measure young children's informal numerical knowledge. The principal limitation of the TEMA is its content validity; it is not a sufficiently broad measure of early mathematical knowledge. All of the items on the TEMA assess number and arithmetic operations, but there are no items that assess children's knowledge of spatial, geometric, or measurement concepts. Therefore, we developed the Child Math Assessment to address the need for a broad measure of children's informal mathematical knowledge.

The full version of the Child Math Assessment (CMA) consisted of 16 tasks that measured informal mathematical knowledge in the areas of number, arithmetic, space and geometry, measurement, and patterns (see Appendix C, Table 1). These tasks were designed to be appropriate for children from 3 to 5 years of age (preschool to kindergarten). Each task included multiple items that encompassed a range of difficulty on a specific concept or skill. Since the CMA was developed to measure children's *informal* mathematical knowledge, all the tasks use concrete objects to engage young children and reveal their mathematical understandings. The selection and design of the tasks were informed by the research literature on mathematical development. Moreover, the breadth of content on the CMA was consistent with the curriculum recommendations of mathematics educators for the pre-kindergarten level (NCTM, 2000, 2006).

Several studies have shown that the CMA provides a comprehensive and broad assessment of preschool children's informal mathematical knowledge. It has been used successfully with both low-income and middle-income children in a longitudinal study of mathematical development in three countries - China, Japan, and the United States (citation?). In the same study, the CMA also revealed evidence of the SES-related gap in mathematical knowledge prior to kindergarten entry, especially in the United States. In addition, the CMA was used as an outcome measure in a rigorous, pre-kindergarten mathematics intervention project (Klein et al., 2008), and it was found to be quite sensitive to the differential growth in mathematical knowledge between treatment and control children over the intervention year. Together, these findings demonstrate the utility of the CMA as a research instrument to measure young children's informal mathematical knowledge. The principal limitation of the CMA, however, is its administration time. The full version of the CMA is administered in two testing sessions on separate days, and each session takes 20-25 minutes. Thus, it would not be practical to use the full CMA in large studies that include a range of early childhood measures such as language and literacy or social skills in addition to math. What was needed for the research community was a shortened version of the CMA.

Determining the Psychometric Properties of the CMA

Although the full CMA had been used successfully in our prior research, we had limited information about the psychometric properties (other than content validity) of this instrument. In addition, it was evident that reducing the length of the instrument to a single testing session would be more practical from a research perspective. Therefore, as part of this Scale-Up Math project, we proposed to develop a shortened version of the CMA and to conduct a separate study of the psychometric properties of this instrument.

The process of determining the psychometric properties of the CMA involved three phases. First, analyzing several previous data sets that included the full set of CMA tasks, we developed a shortened CMA instrument. Second, we conducted a study of this shortened version of the CMA to obtain reliability and validity estimates for the instrument. This phase required collecting new data on the shortened CMA with both low-income and middle-income samples of preschool children in two states during Year 1 of the Scale-Up project. If the shortened CMA were found to be reliable and valid in this study, it would be used as one of the principal math outcome measures (in addition to the TEMA) in the pre-kindergarten years of the Main Study and the Sustainability Study. Finally, we conducted additional Classical Item and IRT analyses of the shortened CMA instrument using data collected during the pre-kindergarten year of the Main Study.

Development of the Shortened CMA Instrument

The process of selecting a subset of CMA tasks for a shortened instrument as well as determining the reliability and validity estimates for the shortened CMA was carried out in collaboration with our psychometric consultants on the Scale-Up Math project.[2] They computed coefficient alphas for each of the 16 CMA tasks based on our previous data sets, and made recommendations about which tasks could be eliminated based on the inter-item reliability values (i.e., coefficient alphas) of those tasks. However, the final decision about which tasks to include in the shortened CMA was made by the PIs based on considerations of content validity, developmental range of difficulty, length of administration time in addition to inter-item reliability estimates.

The first step in developing a shortened CMA was to analyze the performance of pre-kindergarten children on all the CMA tasks in four data sets: Cross-Cultural Math project, U.S. Cohort 1 (n = 88); Cross-Cultural Math project, U.S. Cohort 2 (n = 100); Preschool Curriculum Evaluation Research (PCER) project, Cohort 1 (n = 278); and PCER project, Cohort 2 (n = 313). Thus, the total sample included in this analysis of the full CMA was approximately 800 children. Within each data set, the following values were calculated for each CMA task: (1) Cronbach's alpha, (2) means and standard deviations, and (3) correlations between children's mean scores on individual tasks and their composite (overall) scores on the CMA instrument.

Next, based on this psychometric information, we established several criteria for selecting tasks to be included in the shortened CMA. First, all tasks must have good coefficients alphas (inter-item reliability within a task), taking into account information that the number of items comprising a task affects its inter-item reliability. Second, tasks should not exhibit floor or ceiling effects either at the beginning or the end of the pre-kindergarten year. This insures that all tasks will be sensitive to growth on a particular mathematical concept or skill. Third, tasks should be included from each of the conceptual areas on the full CMA (number, arithmetic, space and geometry, measurement and patterns). This criterion addresses the content validity of the CMA as a broad measure of informal mathematical knowledge. Finally, there is the criterion of administration time. Since one goal of the Psychometric Study was to develop a shortened CMA that could be administered to young children in a single testing session of 25 minutes, time estimates for individual tasks were taken into consideration and stop rules within tasks were added as needed. The overarching objective was to measure as many early mathematical abilities as possible within a limited time period.

The final set of 10 tasks included in the shortened version of the CMA is presented in Table 2 of Appendix C. Note that each conceptual area of children's developing mathematical knowledge is represented by one or more tasks. In developing this shortened CMA, we excluded 8 tasks from the full CMA (see below), added a new task (Division), and separated a previous arithmetic task (Two-Set Addition/Subtraction) into two tasks. The 8 tasks that were excluded and the reasons for excluding them were as follows: (1) evidence of ceiling effects at the end of the pre-k year (Quantitative Comparison, One-Set Addition/Subtraction - Objects Visible, Shape Naming); (2) limited developmental range over the pre-k year (Ordinal Number Terms, Creation of a Non-Standard Unit, Pattern Extension); and (3) relatively low alpha coefficients and limited developmental range (Knowledge of Number Order; Addition/Subtraction Story Problems).

Preliminary psychometric analyses of the shortened CMA (based on previous data sets) revealed that estimates of both the coefficient alpha (>0.82) and the Stratified coefficient alpha (>.93) for this instrument in Fall and Spring of the pre-kindergarten year were very high. Furthermore, these analyses indicated that the Stratified alpha estimate for the shortened version of the CMA was quite stable, and would not change much if a task was added or substituted for another task. Therefore, we proceeded to conduct a full study of the shortened CMA in order to determine the psychometric properties of this instrument.

Psychometric Study of the Shortened CMA

The study was conducted with low-income and middle-income pre-kindergarten children in California (CA) and in Tennessee (TN)[3] during Year 1 of the Scale-Up Math project. There were three principal objectives of the study: (1) Obtain evidence of the inter-item reliability (Cronbach's coefficient alpha) of the tasks comprising the shortened CMA as well as the Stratified coefficient alpha for the overall instrument; (2) obtain evidence of the stability of the shortened CMA by computing test-retest reliability between scores at Time 1 and Time 2; and (3) obtain evidence of criterion-related (concurrent) validity of the shortened CMA with the TEMA.

The total sample consisted of 210 pre-kindergarten children attending preschools in CA (N = 100) and in TN (N = 110). Within each state, efforts were made to equally sample low-income and middle-income children as well as males and females. Children were assessed at mid-year, and the mean age of the total sample at first testing was 4.7 years (range: 4.0 – 5.2

years). The number of children in each demographic category by State is given in the table below:

State	Low-Income	Mid-Income	Males	Females	Total N
California	50	50	52	58	100
Tennessee	62	48	47	53	110

The shortened CMA and the TEMA were administered to all children in a fixed order over three testing sessions. The CMA was administered twice (CMA1 and CMA2) approximately two weeks apart in order to calculate the test-retest reliability of the instrument, and then the TEMA was administered at the third session. The mean interval between administrations of the CMA1 and CMA 2 was 14.32 days, and the mean interval between administrations of the CMA2 and the TEMA was 10.06 days. In addition, 10% of the testing sessions were videotaped to insure consistency of administration of the instruments across assessors and states.

CMA Mean Composite scores were computed for each child at Time 1 and Time 2, and a TEMA Raw score was computed at Time 3. Consistent with reported findings in the research literature, the overall performance of this sample of children on both instruments revealed a striking effect for socioeconomic differences. On average, middle-income children scored 60% greater on the CMA than their low-income peers (.40 vs. .64), and this SES-related difference was even more pronounced on the TEMA (10.38 vs. 19.59).

The next step in the data analyses was to establish the psychometric properties of the shortened CMA. Coefficient alphas were calculated for each task and then Stratified coefficient alphas were computed for the total instrument. Stratified coefficient alphas for the CMA at Time 1 and Time 2 were .923 and .924, respectively, yielding an overall Inter-Item Reliability of .92. This high level of inter-item reliability provided evidence of the internal consistency of the shortened CMA. A second measure of reliability, Test-Retest Reliability, examined the stability of CMA scores over time by calculating a correlation between CMA scores at Time 1 and scores at Time 2. This type of reliability was also found to be quite high, $r = .91$ ($p < .001$). Thus, given the high values of both the inter-item and test-retest reliability coefficients, we concluded that the shortened CMA was reliable instrument - a prerequisite to establishing its validity.

The final psychometric value computed in this study was Criterion-Related Validity. The TEMA was selected as the most appropriate measure with which to establish evidence of criterion validity with the CMA, because it is a reliable and standardized measure of children's informal numerical knowledge. However, since the TEMA is not as broad a measure of informal mathematical knowledge as the CMA (e.g., it does not assess spatial or geometric knowledge), it was not expected that scores on the two instruments would be completely consistent. Children were administered the CMA followed by the TEMA in separate testing sessions. To examine criterion-related validity, correlations were computed between children's Mean Composite scores on the shortened CMA and their Raw scores on the TEMA. As predicted, children's scores on the CMA were significantly correlated with their scores on the TEMA ($r = .77$, $p < .001$), but the relationship between the two measures was not identical. Nevertheless, the high correlation with the TEMA provided evidence for the external validity of the shortened CMA. In summary, the findings of reliability and validity obtained in this study support the conclusion that the CMA has very good psychometric properties for children in this age range.

Classical Item and IRT Analyses of the CMA

Based on the Year 1 findings that the shortened CMA was a reliable and valid measure of children's informal mathematical knowledge, we used this instrument as one of the pre-k math outcome measures in both the Main and Sustainability Studies. This enabled us to conduct additional psychometric analyses of the CMA using the large sample of 744 low-income children who participated in the Main Study of the Scale-Up project (see Main Study for a detailed description of the sample). These analyses were done in collaboration with a statistician who had specialized expertise in Item Response Theory.[4]

The first step involved conducting classical item-level and test-level analyses of the CMA using the pretest and posttest data sets of the Main Study. For each data set, analyses were first performed on the 39 items in the shortened CMA, and then corresponding analyses were performed on the 10 tasks comprising the CMA. This report focuses on two statistics that were computed as part of these analyses: (1) proportion correct (item p-value), and (2) internal consistency estimates (KR-20 based on 39 items, and Stratified alpha based on 10 tasks). The data from the item- and test-level analyses are presented in Table 20. The item analyses revealed a good range of difficulty across the 39 items at both the pretest (.02 - .66) and the posttest (.16 - .94). There were no items at ceiling in the fall and only one item (count "3" objects) that approached ceiling in the spring of Pre-K. Furthermore, within each task, there were more and less difficult items. Both of these properties – an absence of ceiling effects and a range of difficulty within tasks – are desirable in an assessment instrument that measures growth in early mathematical knowledge.

Internal consistency analyses provided additional evidence of the reliability of the shortened CMA. At the item level, the Kuder-Richardson 20 (KR-20) coefficient (a special case of Cronbach's alpha) was computed, and the values exceeded the recommended range (.70 - .80) for group-level comparisons at the pretest and posttest. At the task level, Stratified alpha was computed and found to be high at both the pretest (.91) and the posttest (.92). It should also be noted that these Stratified alpha values were comparable to the value calculated in the Psychometric Study (.92) with a different sample of children. Thus, based on the high inter-item reliability estimates obtained in both studies, the shortened CMA instrument meets the psychometric criterion for internal consistency.

Next, an Exploratory Factor Analysis (EFA) was performed on both the pretest and posttest data sets as a prerequisite for conducting an IRT analysis of the CMA. The primary objective of the factor analyses was to determine whether a unidimensional (i.e., one underlying factor) or a multidimensional model (i.e., two or more underlying factors) was most appropriate for this IRT analysis. For each data set, an EFA was conducted first with 39 items and then with the 10 tasks comprising the CMA. All the factor analyses yielded the same result: there was one dominant factor that we characterized as "math understanding" underlying preschool children's performance on the CMA. Therefore, a unidimensional model was chosen for the IRT analysis.

The final step involved comparing two Rasch IRT models to determine which one provided a better fit between expected and observed scores on the CMA, and then conducting the IRT analysis. The pretest data set from the Main Study was examined in this analysis. A one-parameter logistic model was used to fit the 39 items on the CMA, and it was compared with a partial credit model to fit the 10 tasks. Overall, the partial credit model based on the 10 tasks provided a better fit to the data set, because the CMA is comprised of discrete tasks with multiple items per task and, thus, the items are not locally independent of each other.

Figure 20 illustrates the results of the IRT analysis of the CMA using the partial credit model. This graph depicts the Expected Score Curve for Direct Measurement. Children's estimated understanding of the specific math construct underlying this task is plotted on an IRT theta metric along the abscissa, and the expected score (ranging from 0 to 6, one point per item on this task) for a given theta estimate is plotted along the ordinate. The solid line represents the expected curve, the dotted line represents the observed curve, and the overlap between the two lines represents the fit of this Rasch model to the observed scores on this task. As evident in Figure 20, there is a very good fit of the model to the observed data on Direct Measurement except at the lowest end of the distribution. Specifically, a small proportion of children performed slightly better than expected at the low end of distribution (i.e., children with the lowest estimated math understanding), but the rest of the observed scores fit the expected score curve quite well.

The IRT results obtained for Direct Measurement were similar to those obtained for 9 out of the 10 CMA tasks. On all 9 tasks, there was a good fit using the partial credit model between the observed scores and the expected score curve for a given task. The majority of children performed close to the expected score curve, and any deviations occurred primarily at the lower or upper ends of the estimated distribution where few cases were observed. The one exception was Division. Deviations occurred in the intermediate range as well as at the ends of the estimated distribution for this task, indicating a need to revise the difficulty level of the items comprising this task. In conclusion, the IRT analyses support and extend previous psychometric findings that the CMA is a reliable assessment instrument for young children in this age range. Moreover, these analyses provide additional information about which items on this instrument should be targeted for further development.

Summary and Conclusions

A Psychometric Study of the Child Math Assessment (CMA) instrument was conducted as part of the Scale-Up project in order to: (1) develop a shortened version of the CMA that could be administered to children in a single testing session, (2) obtain reliability and validity estimates for this shortened CMA with a new sample of pre-kindergarten children, and (3) examine additional psychometric properties of the shortened CMA by performing Classical item/test-level analyses and IRT analyses of data from the Main Study of this project. First, a shortened version of the CMA was developed that consisted of 10 tasks, with multiple problems per task, and broadly measured children's informal mathematical knowledge in the areas of number, arithmetic, space and geometry, measurement and patterns. The tasks were designed to be appropriate for children from 3 to 5 years of age, and each task encompassed a range of difficulty on a specific concept or skill.

Next, a full study of the shortened CMA was conducted with 210 pre-kindergarten children from low-income and middle-income backgrounds attending preschools in CA and TN. The shortened CMA and the TEMA-3 were administered to all children in a fixed order across three sessions in order to obtain evidence of inter-item reliability (Cronbach's alpha), test-retest reliability, and criterion-related (concurrent) validity of the CMA with the TEMA-3. The findings from this study indicated that the shortened CMA has very good psychometric properties for children in this age range. The inter-item reliability for the total instrument (Stratified alpha) is .92, and the test-retest reliability of CMA scores over a two-week interval (calculated as a correlation) is .91. Moreover, children's scores on the CMA were significantly correlated with their scores on the TEMA-3 ($r = .77$, $p < .001$), providing strong evidence of the concurrent validity of the CMA with another measure of early mathematical knowledge. Finally, classical item- and test-level analyses revealed that there was a good range of item and task difficulty, with no observable ceiling effects, on the shortened CMA. This provides further support for using the CMA to measure growth in young children's mathematical knowledge. Additional IRT analyses of the shortened CMA were conducted using a partial credit Rasch model to examine item difficulty and fit between observed and expected scores for each of the CMA tasks. Overall, there was a good fit between the observed and expected distributions for 9 of the tasks, and any deviations occurred at the ends of the estimated distribution where there were the fewest cases. Taken together, the findings from these three sets of psychometric analyses indicate that the CMA is an internally consistent, stable, and valid measure of young children's informal mathematical knowledge.

General Conclusions about the Scale-Up Project

The professional development model that was employed in the Main Study enabled teachers to implement with high fidelity, to deliver a good dosage in the classroom, and to motivate parents to deliver a fair-to-good dosage at home. Main impact analyses indicate that the *Pre-K Mathematics* intervention was highly effective when implemented at a customary level of scale, with intervention children experiencing gains in mathematical knowledge about twice as great as control children's gains. The intervention was sufficiently robust that children experienced significant and sustained mathematical growth across varied contexts, including Head Start and state preschool programs serving urban, ethnically diverse families in California and predominantly Caucasian, rural families in Kentucky and Indiana. Mediation analyses supported the hypothesis that the causal influence was transmitted, at least in part, by a set of best math practices, including teaching math-focused activities to children in small groups and providing children with scaffolding as needed.

The Sustainability Study findings indicate that Head Start and state-funded public preschool programs and their teachers were able to sustain implementation of the math intervention with high fidelity and curriculum dosage levels similar to those in the Main Study. The intervention produced significant and sustained effects on children's mathematical knowledge. Children who received the math intervention, relative to control children who did not, also developed a better ability to follow verbal directions from adults.

Recommendations for policy makers. Public preschool programs should be required to implement math curricula that are of proven effectiveness. To build capacity in early math, there is a need to provide professional development in early math to program trainers as well as to preschool teachers. It is also recommended that programs forge a closer working relationship with parents, such as having teachers send pre-k math materials home to parents. We believe that these changes are necessary to make economically disadvantaged children ready for elementary school mathematics.

References

Ball, D.L., & Cohen, D. K. (1999). *Instruction, capacity, and improvement.* Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Baroody, A.J. (2004). The developmental bases for early childhood number and operations standards. In D. H. Clements & J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 173-219). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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: Psychology Press.

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting, 5*, 35–45

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.

Borman, G., Hewes, G., Overman, L., & Brown, S. (2002). *Comprehensive school reform and student achievement: A meta-analysis* (Report No. 59). Baltimore: Center for Research on the Education of Students Placed At Risk.

Bowman, B. T., Donovan, M. S., & Burns, M. S. (Eds.). (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

Bryant, D. M., Burchinal, M., Lau, L. B., Sparkling, J. J. (1994). Family and Classroom Correlates of Head Start Children's Developmental Outcomes. *Early Childhood Research Quarterly, 9*, 289-309.

Clements, D. H., & Sarama, J, & DiBiase A.M. (Eds.). (2004). *Engaging young children in mathematics: Standards early childhood mathematics education*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Copley, J. V. (2004). The early childhood collaborative: A professional development model to communicate and implement the standards. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Findings of the 2000 national conference on standards for preschool and kindergarten mathematics education*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Copley, J. V., & Padrón, Y. N. (1999). Preparing teachers of young learners: Professional development of early childhood teachers in mathematics and science. In J. V. Copley (Ed.) *Dialogue on early childhood science, mathematics, and technology education* (pp.117-129). Washington, DC: American Association for the Advancement of Science.

DeFlorio, L. L. (2011). *The Influence of the Home Learning Environment on Preschool Children's Informal Mathematical Development: Variation by Age and Socioeconomic Status.* Unpublished doctoral dissertation. University of California, Berkeley.

Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade*. Available from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002125 [2002].

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.
Elias, M. J., Zins, J. E., Graczyk, P. A., & Weissberg, R. P. (2003). Implementation, Sustainability, and Scaling Up of Social-Emotional and Academic Innovations in Public *Schools. School Psychology Review, 32*(3), 303-319.
Entwisle, D.R., & Alexander, K.L. (1990). Beginning school competence: Minority and majority comparisons. *Child Development, 61,* 454-471.
Farran, D. C. (1990). Effects of intervention with disadvantaged and disabled children: A decade review. In S. J. Meisels & J. P. Shonkoff (Eds.), *Handbook of early childhood intervention* (pp. 501-539). New York: Cambridge University Press.
Farran, D. C., Silveri, B., & Culp, A. (1991). Public preschools and the disadvantaged. In L. Rescorla, M. C. Hyson, & K. Hirsh-Pasek (Eds.), *Academic instruction in early childhood: Challenge or pressure?* (No. 53, pp. 65–73). *New directions for child development,* W. Damon (editor-in-chief).
Fuson, K. C., Carroll, W. M., Drueck, C. V. (2000). *Achievement results for second and third graders using the standards-based curriculum everyday mathematics*. Journal for Research in Mathematics Education, 31*(3), 277-295.*
Geary, D.C. (1994). *Children's mathematical development*. Washington, D.C.: American Psychological Association.
Gelman, R., & Williams, E.M. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In W. Damon (Series Ed.), D. Kuhn, & R.S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2: Cognition, perception and language* (5th ed., pp. 575-630). New York: Wiley.
Gilliam, W.S., Ripple, C. H., Zigler, E. F., & Leiter, V. (2000). Evaluating Child and Family Demonstration Initiatives: Lessons from the Comprehensive Child Development Program. *Early Childhood Research Quarterly, 15*(1), 41-59.
Ginsburg, H.P., & Baroody, A. (2003). Test of Early Mathematics Ability (TEMA-3), Third Edition. Pro-Ed.
Ginsburg, H. P., Klein, A., & Starkey, P. (1998). The development of children's mathematical thinking: Connecting research with practice. In W. Damon (Series Ed.) & I. E. Sigel & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Vol. 4: Child psychology in practice* (5th ed., pp. 401-476). New York: 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, Serial No. 193).
Glenn Commission. (2000). *Before it's too late: A report to the nation from the National Commission on Mathematics and Science Teaching for the 21st Century*: U.S. Departmentof Education.
Graham, T. A., Nash, C., Paul, K. (1997). Young Children's Exposure to Mathematics: The Child Care Context. *Early Childhood Research Quarterly, 25*(1), 31-38.
Griffin, S., Case, R., & Siegler, R.S. (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: Bradford Books MIT Press.
Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H. Brookes Publishing Co.

Holloway, S. D., Rambaud, M. F., Fuller, B., & Eggers-Pierola, C. (1995). What is "appropriate practice" at home and in child care?: Low-income mothers' views on preparing their children for school. *Early Childhood Research Quarterly, 10,* 451–473.
Huberman, M. A., Miles, M. B., (1984). *Innovation up close: How school improvement works.* New York: Plenum Press.
Hughes, M. (1986). *Children and number: Difficulties in learning mathematics*. Oxford: Blackwell.
IEA (2001). *Mathematics benchmarking report: TIMSS 1999—Eighth grade*. International Association for the Evaluation of Educational Achievement. Available: http://timss.bc.edu/timss1999b/mathbench_report/t99bmath_TOC.html.
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., Huttenlocher, J., & Levine, S.C. (1994). Assessing early arithmetic abilities: Effects of verbal and nonverbal response types on the calculation performance of middle- and low-income children. *Learning and Individual Differences, 6*, 413-432.
Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.
Klein, A., & Starkey, P., (with Ramirez, A.). (2002). *Pre-K mathematics curriculum*. Glendale, IL: Scott Foresman.
Klein, A., & Starkey, P. (2004). *Scott Foresman – Addison Wesley Mathematics: Pre-K*. Glenview, IL: Pearson Scott Foresman. [2nd edition of Pre-K Mathematics]
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.
Klinger, J., Ahwee, S., Pilonieta, P., & Menendez, R. (2003). Barriers and facilitators on scaling up research-based practices. Exceptional Children, 69, 411–429.
Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *J Learn Disabil, 42*(4), 306-321.
Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1997). *Mathematics achievement in the primary school years: IEA's third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A.,O'Connor,K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999 international mathematics report*. Boston, MA: The International Study Center, Boston College, Lynch School of Education.
National Association for the Education of Young Children (2010). *Environment Rating Scales and NAEYC Accreditation.* Retrieved from http://www.naeyc.org/files/academy/file/ERSandAccreditationPDI2010.pdf
National Council of Teachers of Mathematics (2000*). Principles and standards for school mathematics*. Reston, VA: NCTM.
National Mathematics Advisory Panel. (2008). *Final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.
Natriello, G., McDill, E. L., & Pallas, A. M. (1990). *Schooling disadvantaged children:*

Racing against catastrophe. New York: Teachers College Press.
Piaget, J. (1952). *The child's conception of number*. London: Routledge & Kegan Paul Ltd.
Preschool Curriculum Evaluation Research Consortium (2008). *Effects of preschool curriculum programs on school readiness.* Washington, D.C., U.S. Department of Education.
Rathbun, A., & West, J. (2004). *From kindergarten through third grade: Children's beginning school experience.* National Center for Education Statistics. Available from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2--4--7#
Resnick, G., Zill, N. (1999). Is Head Start providing high-quality educational services? "Unpacking" classroom processes. Presented at the meeting of the Society for Research in Child Development, Albuquerque, New Mexico.
Riordan, J. E., Noyce, P. E. (2001). *The impact of two standards-based mathematics curricula on student achievement in Massachusetts.* Journal for Research in Mathematics Education,32*(4), 368-398.*
Saxe, G. B., Guberman, S. R., & Gearhart, M. (1987). Social processes in early number development. *Monographs of the Society for Research in Child Development, 52*, (2 Serial No. 216).
Shulman, L. S. (2000). Teacher development: Roles of domain expertise and pedagogical knowledge. *Journal of Applied Developmental Psychology 21*(1): 129-135.
Sophian, C. (1996). *Children's numbers*. Madison, WI: Brown & Benchmark.
Starkey, P. (1992). The early development of numerical reasoning. *Cognition, 43*, 93-126.
Starkey, P. (2003). *Young children's mathematical development and learning environments in China, Japan, and the United States*. Symposium conducted at the meeting of the Society for Research in Child Development, Tampa, Florida.
Starkey, P., & Cooper, R.G. (1980). Perception of numbers by human infants. *Science*, *210*, 1033-1035.
Starkey, P., & Klein, A. (1992). Economic and cultural influences on early mathematical development. In F. L. Parker, R. Robinson, S. Sombrano, C. Piotrowski, J. Hagen, S. Randolph, and A. Baker (Eds.), *New directions in child and family research: Shaping Head Start in the 90s* (p. 440). New York: National Council of Jewish Women.
Starkey, P., & Klein, A. (2000). Fostering parental support for children's mathematical development: An intervention with Head Start families. *Early Education and Development, 11*, 659-680.
Starkey, P., & Klein, A. (2003). *Supporting Young Children's Readiness for School Mathematics Through a Pre-kindergarten Classroom and Family Math Curriculum*. Final Report: OERI/US Department of Education Grant R307F60024.
Starkey, P., & Klein, A. (2008). Sociocultural influences on young children's mathematical knowledge. In Saracho, O. N. and Spodek, B (Eds.) *Contemporary perspectives on mathematics in early childhood education*. (pp.253-276). Charlotte, NC: Information Age Publishing
Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly, 19*(1), 99-120.
Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36,* 97-127.
Super, C.M., & Harkness, S. (1996). The cultural structuring of child development. In J.W. Berry, P.R. Dasen, & T.S. Saraswathi (Eds.), *Handbook of cross-cultural psychology*

Vol.2: Basic processes and human development (2nd ed., pp. 1-39). Boston: Allyn & Bacon.
U.S. Department of Health and Human Services, Administration for Children and Families (2010). Head Start Impact Study. Final Report. Washington, DC.
U.S. Department of Labor Bureau of Labor Statistics. (2000, Spring). The Outlook for College Graduates, 1998–2008. In *Getting ready pays off!*, U.S. DOE, October 2000, and BLS Occupational Employment Projections to 2008, in NAB, *Workforce Economics, 6*(1).
Vygotsky, L.S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
Wakeley, R.A. (2002). *Early mathematical development in very low birthweight children*. Unpublished doctoral dissertation, University of California, Berkeley.
West, J., Denton, K., & Germino-Hausken, E. (2000). *America's kindergarteners*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
Woodock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement* (3rd ed). Itasca, IL: Riverside Publishing Company.
Zill, N., Resnick, G., Kim, K., McKey, R. H., Clark, C., Pai-Samant, S., D'Elio, M. A. (2001). *Head Start FACES: Longitudinal findings on program performance. Third progress report*. Washington, DC: U.S. Department of Health and Human Services, Administration on Children, Youth, and Families.

Appendix A
Tables and Figures

<u>Table of Contents</u>

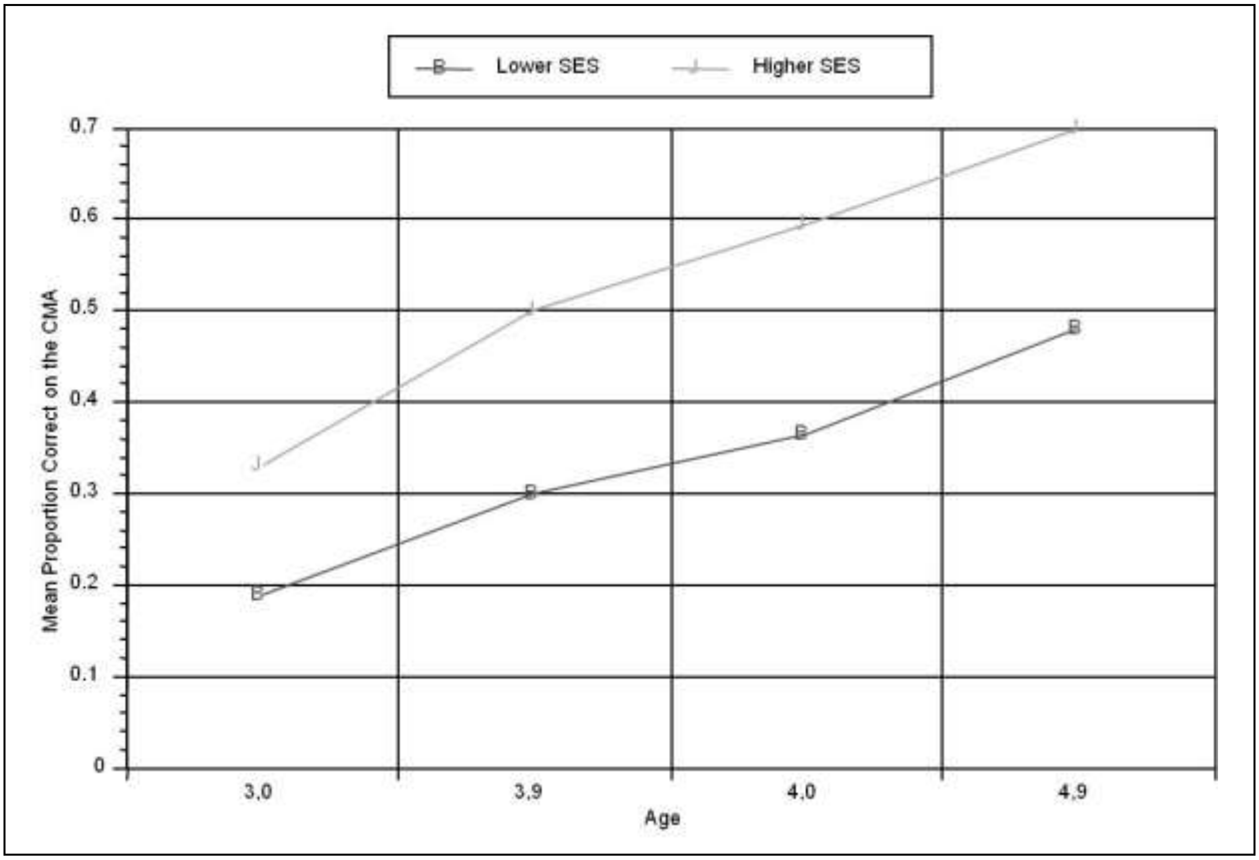

Figure 1. CMA scores of American children.

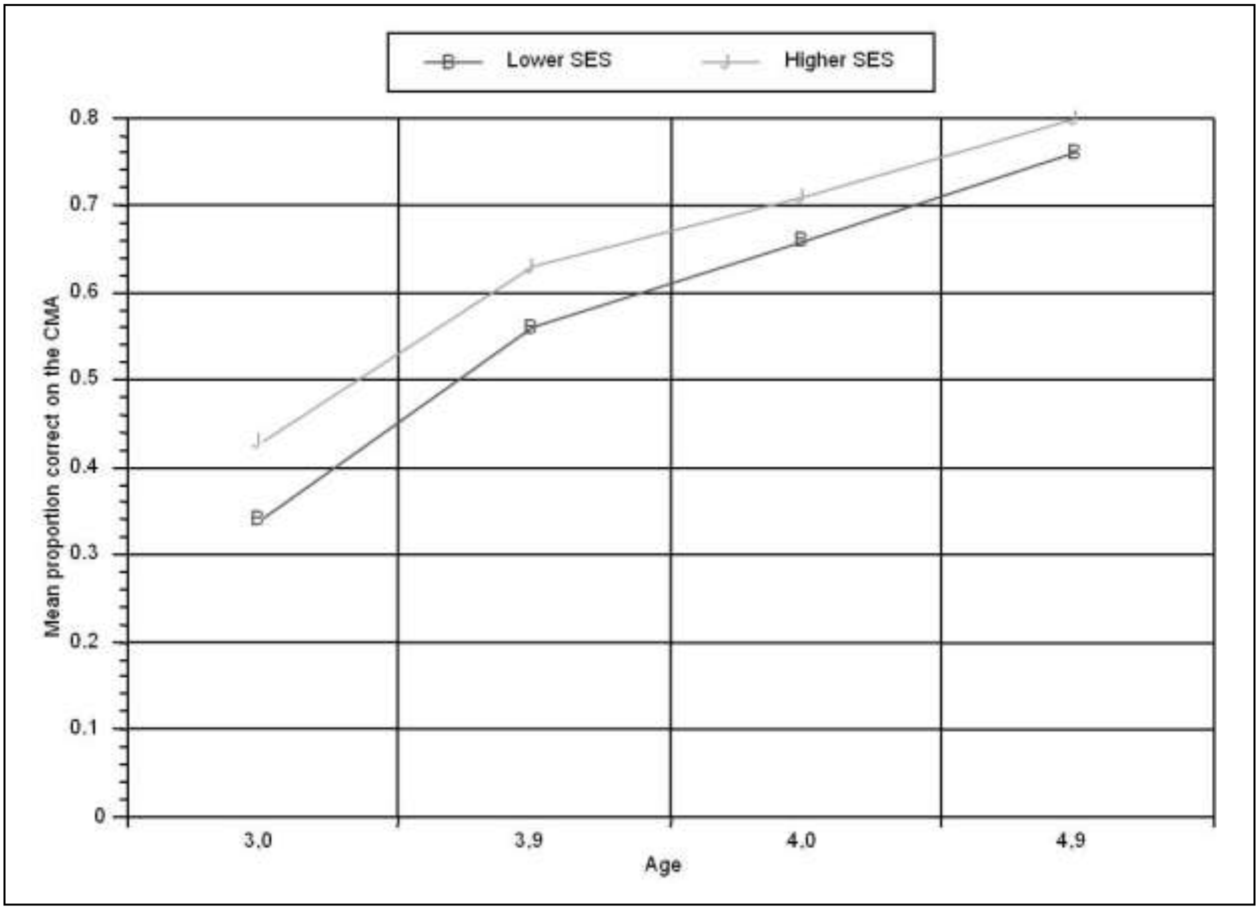

Figure 2. CMA scores of Chinese children.

Table 1
Main Study Child Demographics

	Age			Gender				Language			
				Male		Female		English		Spanish	
	n	*M*	*Range*	*n*	(%)	*n*	(%)	*n*	(%)	*n*	(%)
Total Sample	744	4.44	3.86 – 4.99	343	(46.1%)	401	(53.9%)	672	(90.3%)	72	(9.7%)
Treatment	387	4.43	3.87 – 4.99	187	(48.3%)	200	(51.7%)	362	(93.5%)	25	(6.5%)
CA	181	4.39	3.87 – 4.80	86	(47.5%)	95	(52.5%)	156	(86.2%)	25	(13.8%)
KY	206	4.47	3.93 – 4.99	101	(49.0%)	105	(51.0%)	206	(100.0%)	0	(0.0%)
Head Start	208	4.43	3.87 – 4.99	100	(48.1%)	108	(51.9%)	190	(91.3%)	18	(8.7%)
State Preschool	179	4.43	3.87 – 4.92	87	(48.6%)	92	(51.4%)	172	(96.1%)	7	(3.9%)
CA Head Start	99	4.37	3.87 – 4.80	45	(45.5%)	54	(54.5%)	81	(81.8%)	18	(18.2%)
KY Head Start	109	4.49	3.94 – 4.99	55	(50.5%)	54	(49.5%)	109	(100.0%)	0	(0.0%)
CA State Preschool	82	4.41	3.87 – 4.79	41	(50.0%)	41	(50.0%)	75	(91.5%)	7	(8.5%)
KY State Preschool	97	4.44	3.93 – 4.92	46	(47.4%)	51	(52.6%)	97	(100.0%)	0	(0.0%)
Control	357	4.45	3.86 – 4.99	156	(43.7%)	201	(56.3%)	310	(86.8%)	47	(13.2%)
CA	186	4.37	3.86 – 4.85	82	(44.1%)	104	(55.9%)	139	(74.7%)	47	(25.3%)
KY	171	4.53	4.01 – 4.99	74	(43.3%)	97	(56.7%)	171	(100.0%)	0	(0.0%)
Head Start	196	4.44	3.86 – 4.99	82	(41.8%)	114	(58.2%)	166	(84.7%)	30	(15.3%)
State Preschool	161	4.45	3.89 – 4.99	74	(46.0%)	87	(54.0%)	144	(89.4%)	17	(10.6%)
CA Head Start	102	4.34	3.86 – 4.85	40	(39.2%)	62	(60.8%)	72	(70.6%)	30	(29.4%)
KY Head Start	94	4.55	4.06 – 4.99	42	(44.7%)	52	(55.3%)	94	(100.0%)	0	(0.0%)
CA State Preschool	84	4.40	3.89 – 4.85	42	(50.0%)	42	(50.0%)	67	(79.8%)	17	(20.2%)
KY State Preschool	77	4.50	4.01 – 4.99	32	(41.6%)	45	(58.4%)	77	(100.0%)	0	(0.0%)

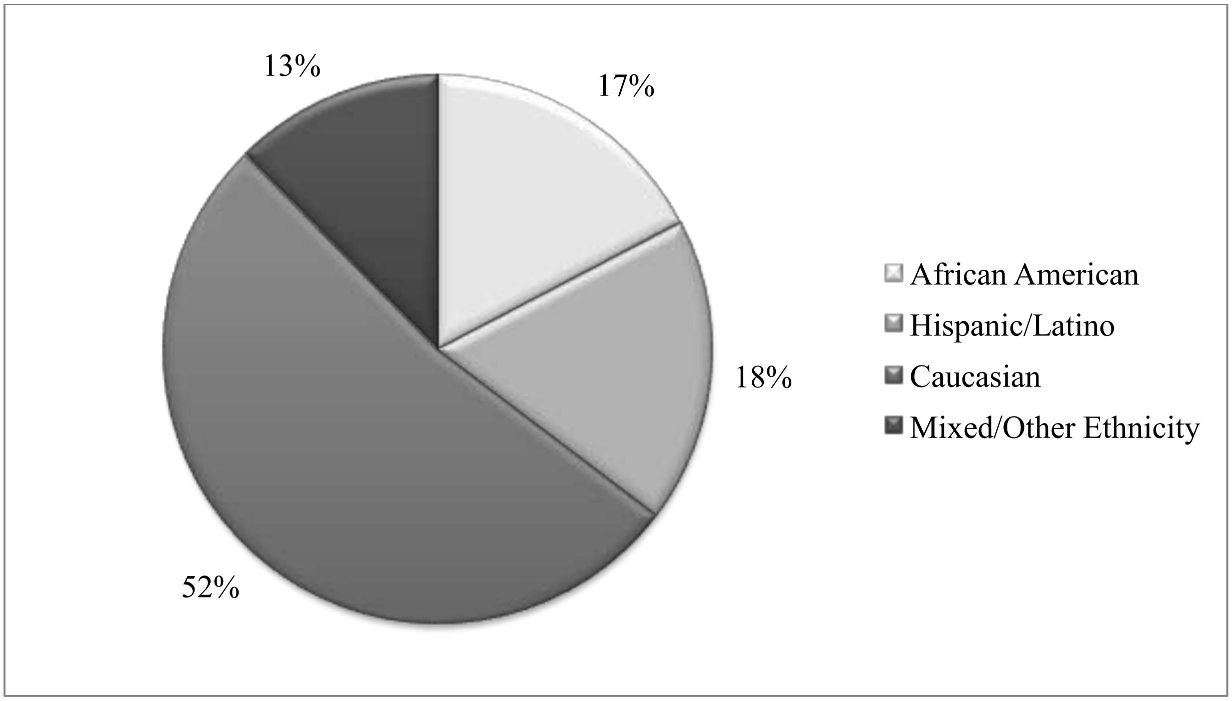

Figure 3. Main Study child ethnicity.

Table 2

Main Study: Fidelity of Implementation Scores

	California		Kentucky/Indiana		All Treatment
Dimension of Fidelity	State PreK	Head Start	State PreK	Head Start	Teachers
Teacher follows a schedule	0.92	0.98	0.88	0.93	.93
Teacher is prepared	0.96	0.92	0.88	0.97	.93
Teacher delivers basic activity correctly	0.95	0.96	0.86	0.96	.93
Teacher makes appropriate developmental adjustments	0.78	0.92	0.85	0.93	.88
Teacher keeps written records about each child's performance	0.86	0.90	0.80	0.89	.86
Overall fidelity	**0.89**	**0.93**	**0.86**	**0.93**	**.91**

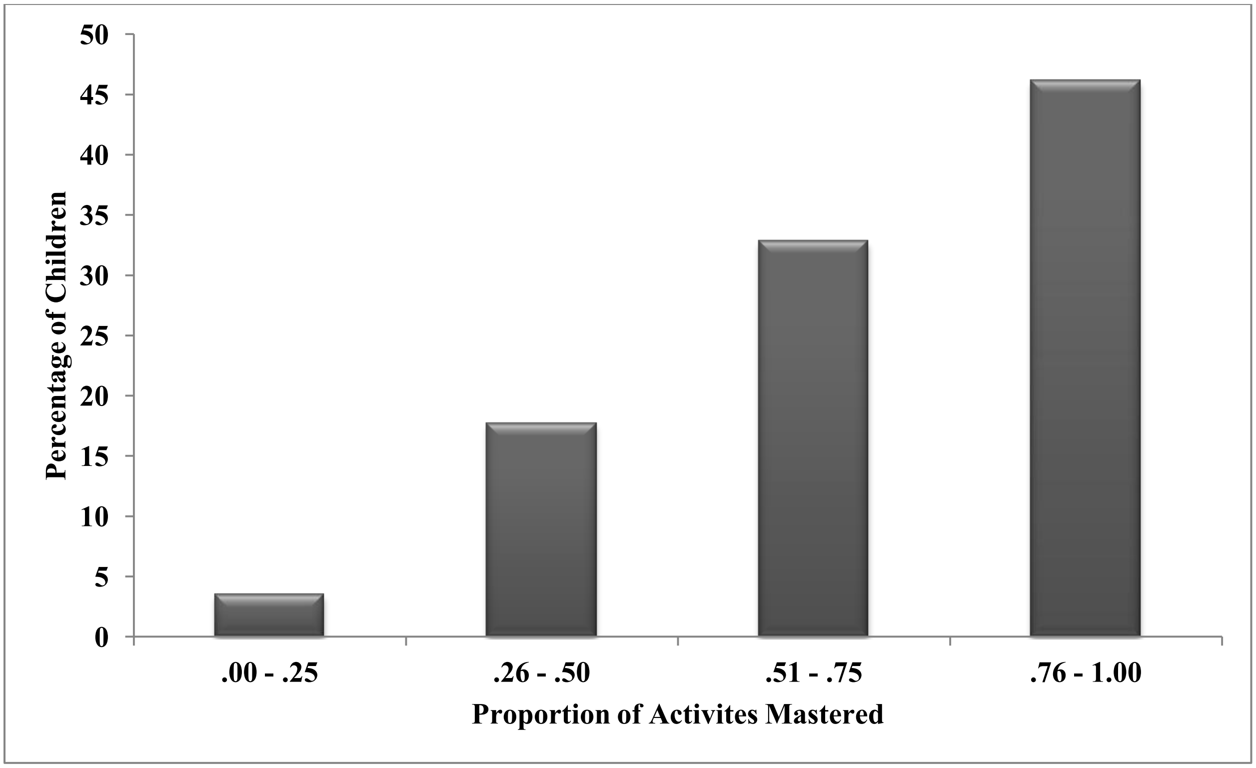

Figure 4. Percentage of children in the Main Study mastering *Pre-K Mathematics* activities. Controlling for pretest scores, regression analyses revealed that mastery of the math curriculum activities significantly predicted posttest scores on both the CMA, F (2, 342) = 161.24, p < .0001, and the TEMA-3, F (2, 342) = 294.04, p < .0001.

Table 3

Main Study: Mean Scores (and Standard Deviations) on the CMA and TEMA-3 in the Pre-kindergarten Year

	CMA				TEMA-3			
	Pretest		Posttest		Pretest		Posttest	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Treatment	.29	(.17)	.61	(.19)	7.02	(4.94)	14.82	(6.95)
CA	.28	(.16)	.61	(.19)	7.01	(5.26)	14.65	(7.16)
KY	.30	(.18)	.61	(.19)	7.03	(4.66)	14.95	(6.77)
Head Start	.27	(.17)	.58	(.19)	6.79	(4.99)	14.29	(6.46)
State PK	.32	(.18)	.64	(.18)	7.29	(4.89)	15.40	(7.42)
CA Head Start	.25	(.15)	.57	(.19)	6.65	(5.27)	13.91	(5.91)
KY Head Start	.29	(.18)	.60	(.19)	6.92	(4.74)	14.59	(6.89)
CA State PK	.33	(.17)	.65	(.18)	7.44	(5.25)	15.42	(8.23)
KY State PK	.31	(.18)	.62	(.19)	7.17	(4.59)	15.37	(6.66)
Control	.32	(.17)	.48	(.19)	7.05	(5.07)	12.49	(6.64)
CA	.31	(.16)	.47	(.17)	7.02	(5.06)	12.34	(6.03)
KY	.33	(.19)	.50	(.20)	7.08	(5.10)	12.66	(7.27)
Head Start	.32	(.17)	.49	(.19)	7.17	(5.00)	12.51	(6.70)
State PK	.31	(.18)	.48	(.18)	6.90	(5.17)	12.47	(6.58)
CA Head Start	.30	(.15)	.45	(.18)	6.67	(4.58)	11.57	(5.63)
KY Head Start	.35	(.18)	.53	(.20)	7.72	(5.39)	13.47	(7.55)
CA State PK	.32	(.17)	.50	(.17)	7.45	(5.58)	13.21	(6.38)
KY State PK	.30	(.19)	.45	(.19)	6.30	(4.65)	11.55	(6.76)

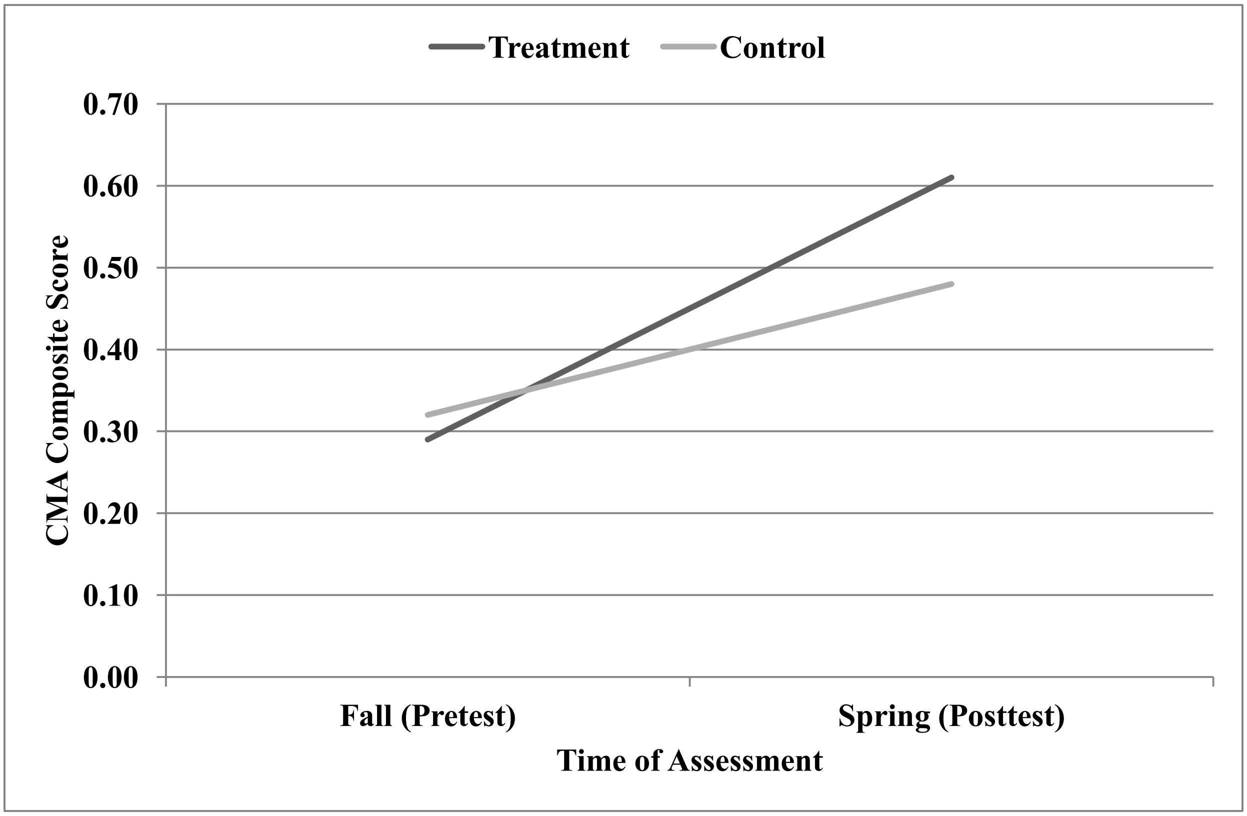

Figure 5. The effects of the Main Study intervention on children's mathematical knowledge in pre-kindergarten as measured by the CMA.

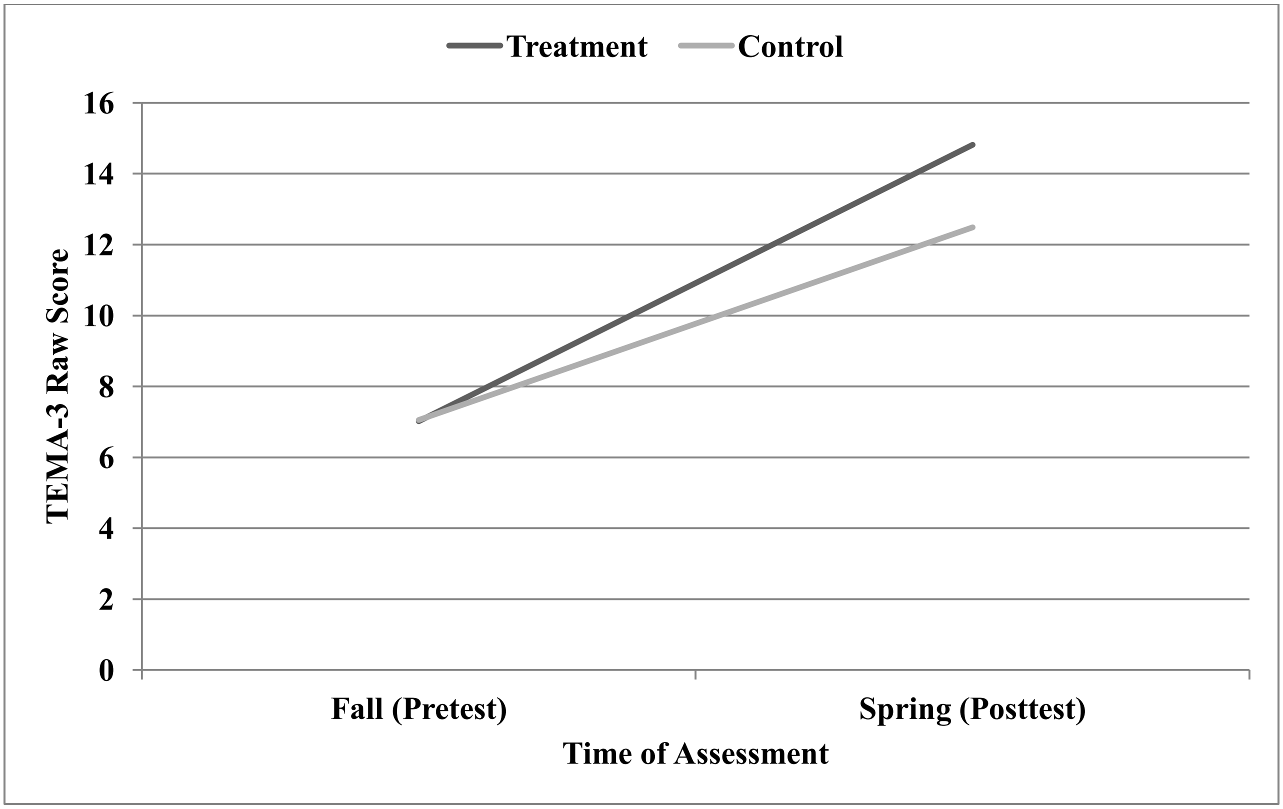

Figure 6. The effects of the Main Study intervention on children's mathematical knowledge in pre-kindergarten as measured by the TEMA-3.

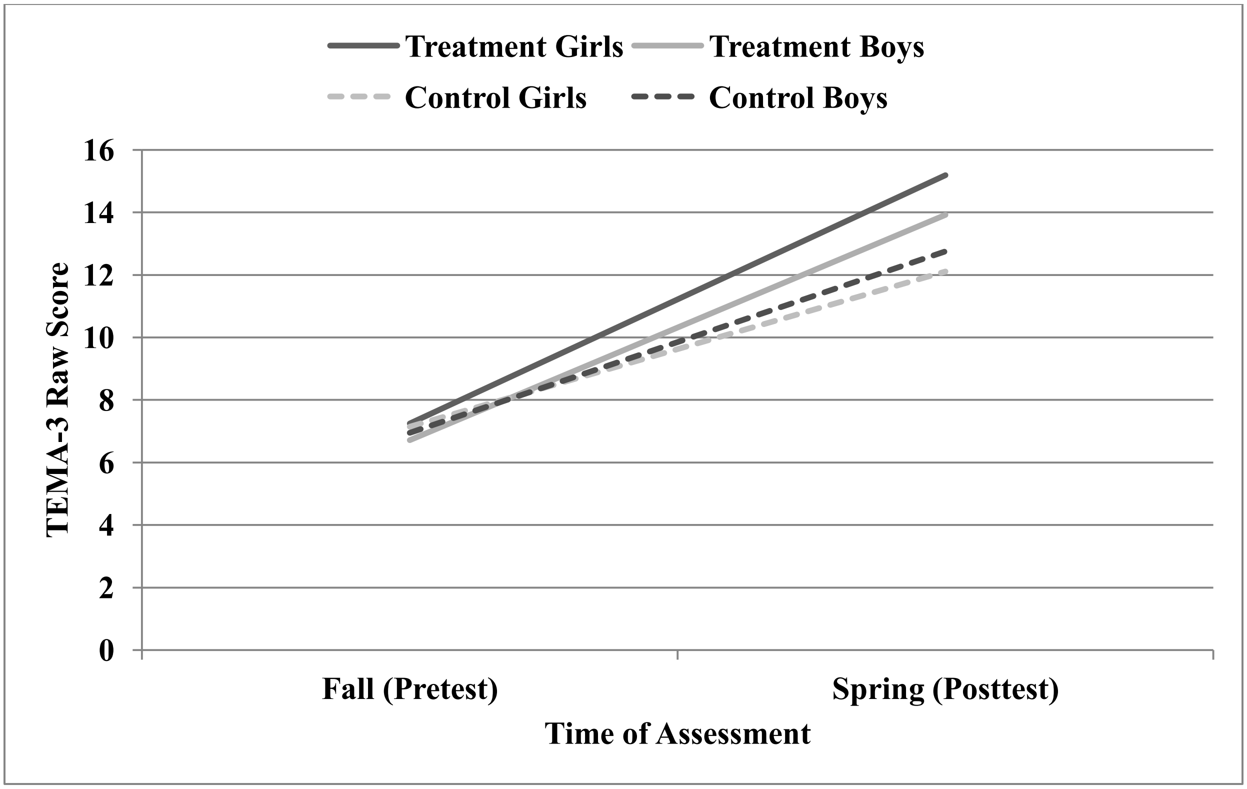

Figure 7. The effects of the Main Study intervention on children's mathematical knowledge in pre-kindergarten by time, condition, and gender.

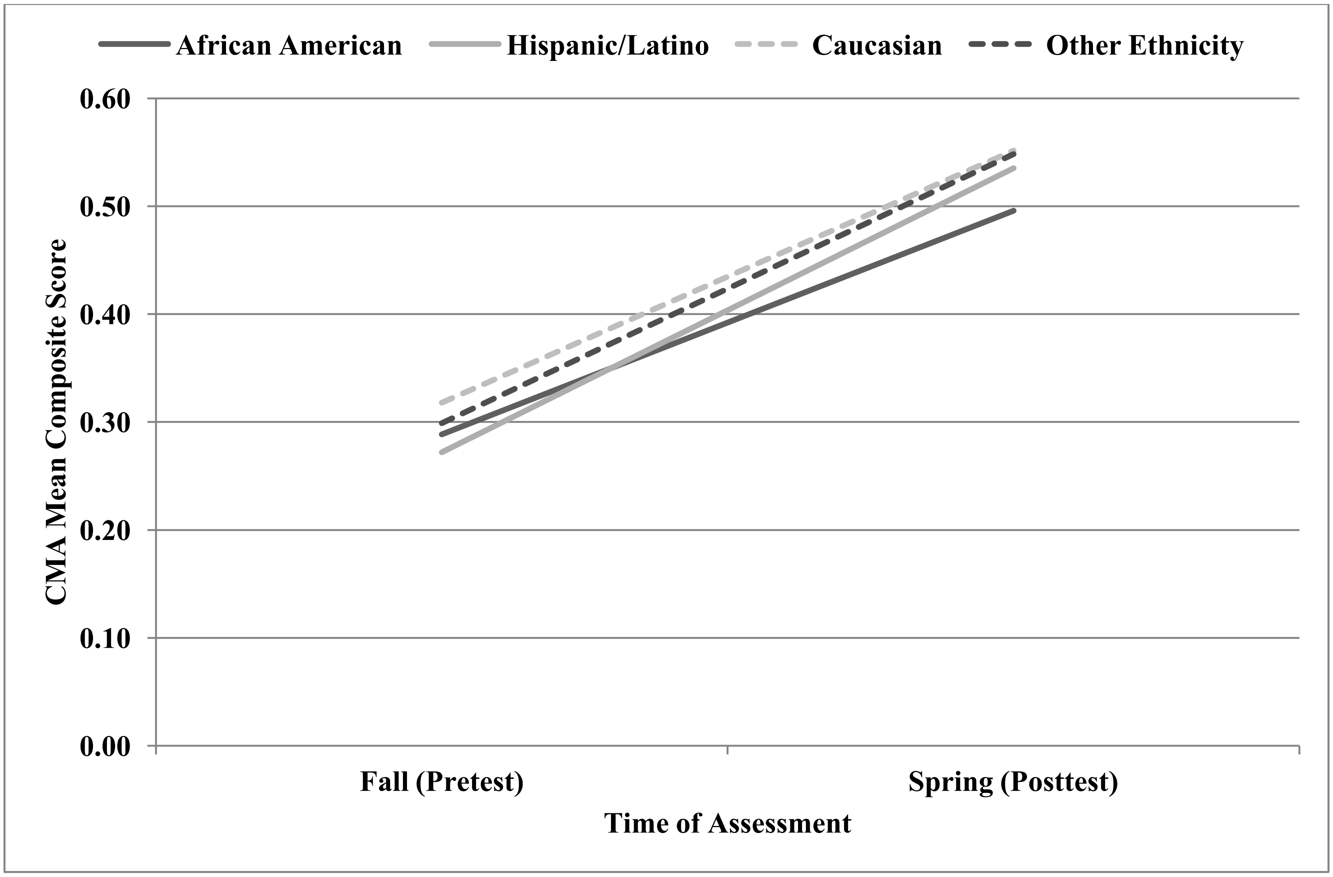

Figure 8. The effects of the Main Study intervention on children's mathematical knowledge in pre-kindergarten by time and ethnicity.

Table 4

Main Study: Mean Scores (and Standard Deviations) on the WJ-III Subtests in the Pre-kindergarten Year

	Understanding Directions				Letter-Word Identification				Spelling			
	Pretest		Posttest		Pretest		Posttest		Pretest		Posttest	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Treatment	451.72	(12.66)	456.04	(11.47)	320.80	(22.18)	340.65	(22.34)	358.57	(23.35)	379.30	(22.34)
CA	448.93	(12.94)	455.95	(11.20)	323.45	(22.38)	339.66	(22.23)	356.41	(25.72)	375.72	(21.94)
KY	454.18	(11.91)	456.12	(11.73)	318.48	(21.79)	341.50	(22.46)	360.47	(20.93)	382.35	(22.28)
Head Start	450.69	(12.72)	454.25	(11.25)	320.41	(22.44)	339.51	(22.34)	356.92	(23.46)	378.00	(22.77)
State PK	452.93	(12.51)	458.04	(11.42)	321.26	(21.93)	341.93	(22.34)	360.49	(23.14)	380.74	(21.83)
CA Head Start	447.31	(13.43)	453.01	(10.78)	321.51	(23.23)	336.43	(21.81)	354.37	(24.95)	375.16	(22.20)
KY Head Start	453.75	(11.24)	455.26	(11.58)	319.41	(21.75)	342.00	(22.57)	359.24	(21.87)	380.30	(23.07)
CA State PK	450.88	(12.10)	459.04	(10.87)	325.79	(21.21)	343.05	(22.31)	358.87	(26.56)	376.31	(21.79)
KY State PK	454.66	(12.66)	457.13	(11.89)	317.42	(21.90)	340.92	(22.45)	361.86	(19.83)	384.75	(21.20)
Control	450.37	(12.67)	455.28	(11.43)	320.78	(22.48)	339.52	(22.32)	360.92	(23.41)	378.30	(23.98)
CA	447.76	(12.81)	453.46	(10.99)	323.36	(21.88)	337.23	(20.84)	359.32	(24.68)	374.74	(23.74)
KY	453.20	(11.92)	457.27	(11.60)	317.97	(22.85)	342.02	(23.64)	362.67	(21.89)	382.19	(23.71)
Head Start	449.85	(12.37)	455.05	(11.78)	322.01	(22.57)	340.55	(22.22)	361.65	(23.81)	378.28	(23.58)
State PK	451.00	(13.03)	455.58	(11.03)	319.29	(22.35)	388.26	(22.45)	360.04	(22.96)	378.33	(24.54)
CA Head Start	446.76	(11.93)	452.20	(11.04)	323.29	(22.03)	335.12	(21.45)	359.99	(24.71)	373.09	(23.19)
KY Head Start	453.20	(12.02)	457.89	(11.86)	320.61	(23.17)	345.97	(21.76)	363.46	(22.79)	383.46	(22.95)
CA State PK	448.99	(13.77)	454.87	(10.83)	323.44	(21.82)	339.61	(19.99)	358.50	(24.76)	376.60	(24.35)
KY State PK	453.20	(11.87)	456.43	(11.29)	314.75	(22.17)	336.62	(25.18)	361.71	(20.85)	380.45	(24.79)

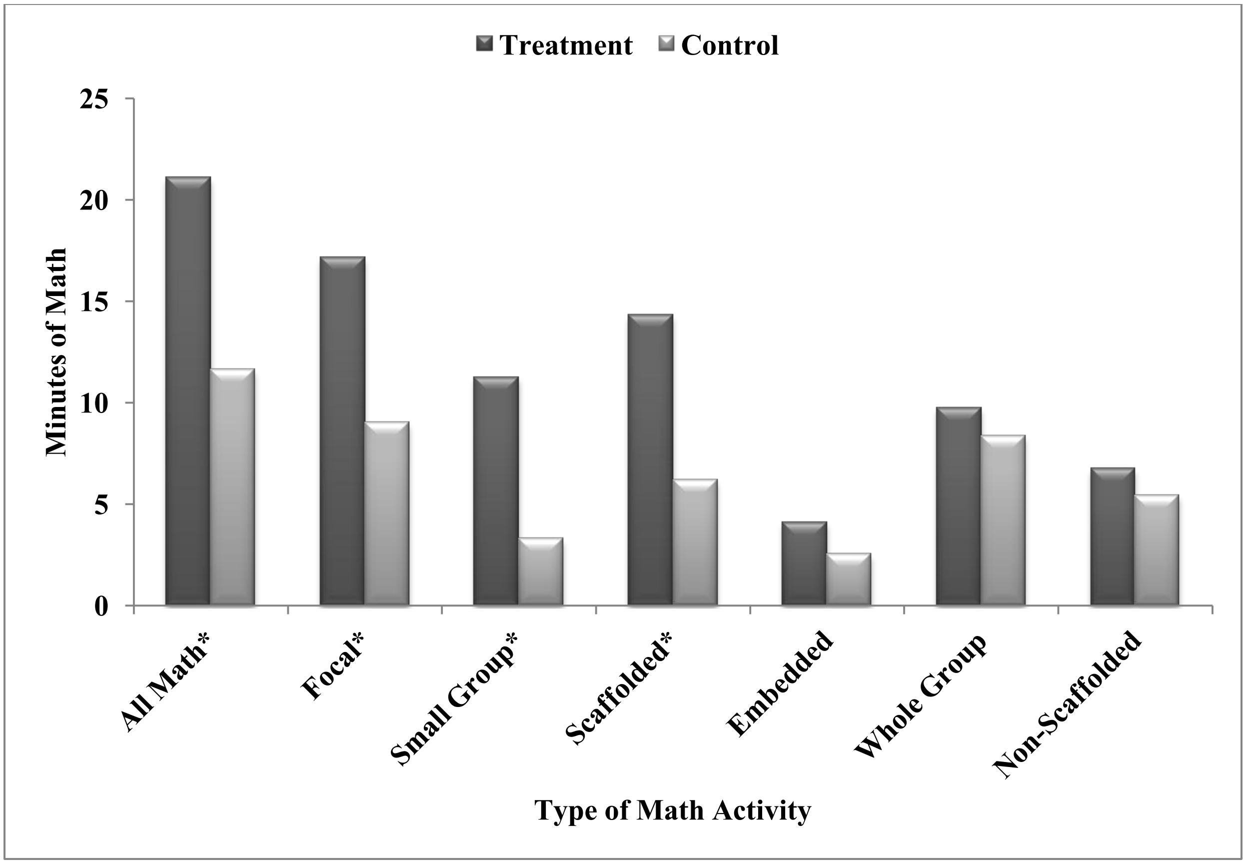

Figure 9. Math practices by pre-kindergarten treatment and control teachers in the Main Study. Significant differences are denoted by *, $p < .0001$.

Mediation Model

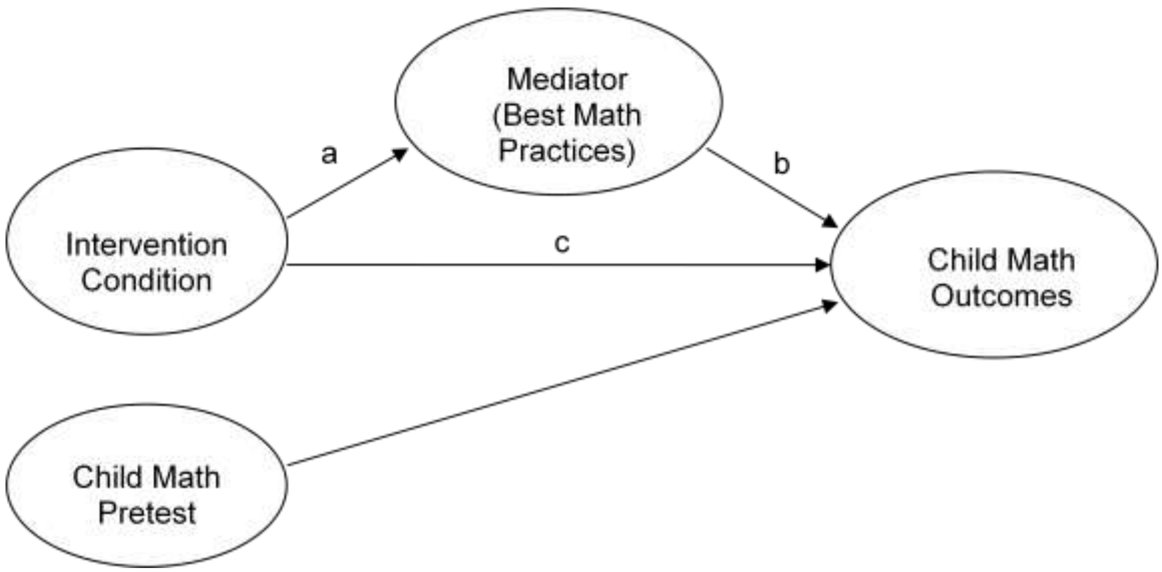

Figure 10. Mediation model.

Table 5

Classroom Mediator Variables that Predict Growth in Children's Mathematical Knowledge

	CMA Outcomes		TEMA-3 Outcomes	
	t	*p*	*t*	*p*
Total MOM	4.82	< .0001	4.54	< .0001
Small-Group MOM	5.68	< .0001	4.54	< .0001
Focal MOM	5.00	< .0001	4.36	< .0001
Scaffolded MOM	4.56	< .0001	4.35	< .0001
Small-Group, Focal, Scaffolded MOM	6.12	< .0001	4.63	< .0001

Table 6

Classroom Variables that Do Not Predict Growth in Children's Mathematical Knowledge

	CMA Outcomes		TEMA-3 Outcomes	
	t	*p*	*t*	*p*
Whole-Group MOM	1.01	ns	1.85	ns
Embedded MOM	1.08	ns	1.68	ns
Non-Scaffolded MOM	1.85	ns	1.65	ns

Table 7
Effects of the Intervention on TEMA-3 Scores with the Total Minutes of Math (TOTMOM) as the Mediator.

Effect	DV	Mediator	*p* =.05	*p* =.25	*p* =.50	*p* =.95	*p* =.975
Indirect	TEMA-3	TOTMOM	-0.77316	-0.82624	-0.47740	-0.19342	-0.14377
A Parameter	TEMA-3	TOTMOM	-8.96231	-9.00094	-8.77156	-8.48947	-8.36595
B Parameter	TEMA-3	TOTMOM	0.022283	0.016376	0.054862	0.08848	0.09449
C Parameter	TEMA-3	TOTMOM	-2.55491	-2.72206	-1.68147	-0.80349	-0.62952

Variable	*N*	*M*	*SD*	Minimum	Maximum
Indirect	5000	-0.4806002	0.1737612	-1.0870496	0.0858782
A Estimate	5000	-8.7277178	0.1512216	-9.3737334	-8.0104960
B Estimate	5000	0.0550709	0.0198960	-0.0097905	0.1268008
C Estimate	5000	-1.6797805	0.5300406	-3.3943536	0.4057238

Table 8
Effects of the Intervention on TEMA-3 Scores with the Small Group Minutes of Math (SGMOM) as the Mediator.

Effect	DV	Mediator	*p* =.05	*p* =.25	*p* =.50	*p* =.95	*p* =.975
Indirect	TEMA-3	SGMOM	-1.10343	-1.20216	-0.62663	-0.14985	-0.06488
A Parameter	TEMA-3	SGMOM	-8.18308	-8.18308	-8.06129	-7.77770	-7.70159
B Parameter	TEMA-3	SGMOM	0.018787	0.007996	0.077866	0.13729	0.14993
C Parameter	TEMA-3	SGMOM	-2.54339	-2.71469	-1.60000	-0.66829	-0.49285

Variable	*N*	*M*	*SD*	Minimum	Maximum
Indirect	5000	-0.6268666	0.2897505	-1.7679783	0.3987608
A Estimate	5000	-8.0124406	0.1240031	-8.4122614	-7.5226760
B Estimate	5000	0.0782182	0.0361259	-0.0498738	0.2193171
C Estimate	5000	-1.6053830	0.5677000	-3.7733920	0.3584838

Table 9
Effects of the Intervention on TEMA-3 Scores with the Focal Minutes of Math (FOCMOM) as the Mediator.

Effect	DV	Mediator	*p* =.05	*p* =.25	*p* =.50	*p* =.95	*p* =.975
Indirect	TEMA-3	FOCMOM	-0.80913	-0.88365	-0.48453	-0.16042	-0.10039
A Parameter	TEMA-3	FOCMOM	-7.97697	-8.00027	-7.81725	-7.55298	-7.48647
B Parameter	TEMA-3	FOCMOM	0.020523	0.012923	0.062287	0.10441	0.11350
C Parameter	TEMA-3	FOCMOM	-2.58145	-2.74500	-1.70091	-0.81070	-0.63777

Variable	*N*	*M*	*SD*	Minimum	Maximum
Indirect	5000	-0.4856018	0.1976848	-1.3207210	0.2338121
A Estimate	5000	-7.7752366	0.1348627	-8.3950247	-7.0822674
B Estimate	5000	0.0624441	0.0253718	-0.0296231	0.1689495
C Estimate	5000	-1.6930935	0.5397008	-3.7340190	0.4859405

Table 10
Effects of the Intervention on TEMA-3 Scores with the Scaffolded Minutes of Math (SCFMOM) as the Mediator.

Effect	DV	Mediator	*p* =.05	*p* =.25	*p* =.50	*p* =.95	*p* =.975
Indirect	TEMA-3	SCFMOM	-2.52174	-2.68055	-1.61488	-0.68530	-0.52699
A Parameter	TEMA-3	SCFMOM	-7.86216	-7.94507	-7.77491	-7.48072	-7.39427
B Parameter	TEMA-3	SCFMOM	0.029012	0.021542	0.067405	0.10670	0.11490
C Parameter	TEMA-3	SCFMOM	-0.82323	-0.89158	-0.52031	-0.22298	-0.16664

Variable	*N*	*M*	*SD*	Minimum	Maximum
Indirect	5000	-0.5241235	0.1820034	-1.1660741	0.1321500
A Estimate	5000	-7.7214645	0.1315591	-8.3375028	-7.1960272
B Estimate	5000	0.0678695	0.0235305	-0.0177265	0.1500069
C Estimate	5000	-1.6129463	0.5571647	-3.6423397	0.6551524

Table 11

Main Study Treatment Group: Mean Number of Activities Parents Reported Children Doing at Home in 1 Week

Activity	*M*	*SD*
Book and Language Activities		
Heard or read/looked at a story book	5.09	4.73
Heard or read/looked at a counting or shape book*	2.43	3.24
Heard or read/looked at a children's magazine or children's section of a newspaper	1.09	2.09
Heard or read/looked at other types of books	2.63	3.79
Heard or told children's stories	2.66	3.82
Heard or sang children's songs	4.77	5.68
Engaged in real or pretend writing activities	3.16	4.18
Store-Bought Games		
Played a board game	1.88	2.82
Played a card game	1.72	2.64
Activities Using Technology		
Watched an educational TV program or video	4.73	5.10
Played an electronic game or video game with educational content	2.81	3.92
Used a computer with children's educational software or visited a children's website with educational content	1.94	3.17
Used a computer with children's math software or visited a children's website with math content*	.99	2.17
Educational Activities or Toys		
Played with wooden or plastic blocks	2.04	3.12
Played with puzzles, shape tiles, mazes, or other spatial materials*	2.79	3.37
Engaged in origami (paper-folding) or kitagami (paper cutting) activities*	1.31	2.28
Used materials that teach letters to children	2.78	3.71
Used materials that teach numerals to children*	2.44	3.71
Home Routine		
Helped set the table	2.49	3.01
Helped prepare food	1.89	2.56
Helped sort laundry	1.95	2.65
Activities Away from Home (Excluding Preschool)		
Attended a supplementary enrichment program for young children	.67	1.94
Went shopping with an adult	1.75	2.17
*Focal Math Activities	9.95	10.09

Table 12

Main Study: Mean Scores (and Standard Deviations) on the TEMA-3 in Kindergarten and Grade 1

	Kindergarten TEMA-3		Grade 1 TEMA-3	
	M	*(SD)*	*M*	*(SD)*
Treatment	26.63	(8.96)	41.81	(8.66)
Control	25.67	(8.42)	41.08	(9.10)

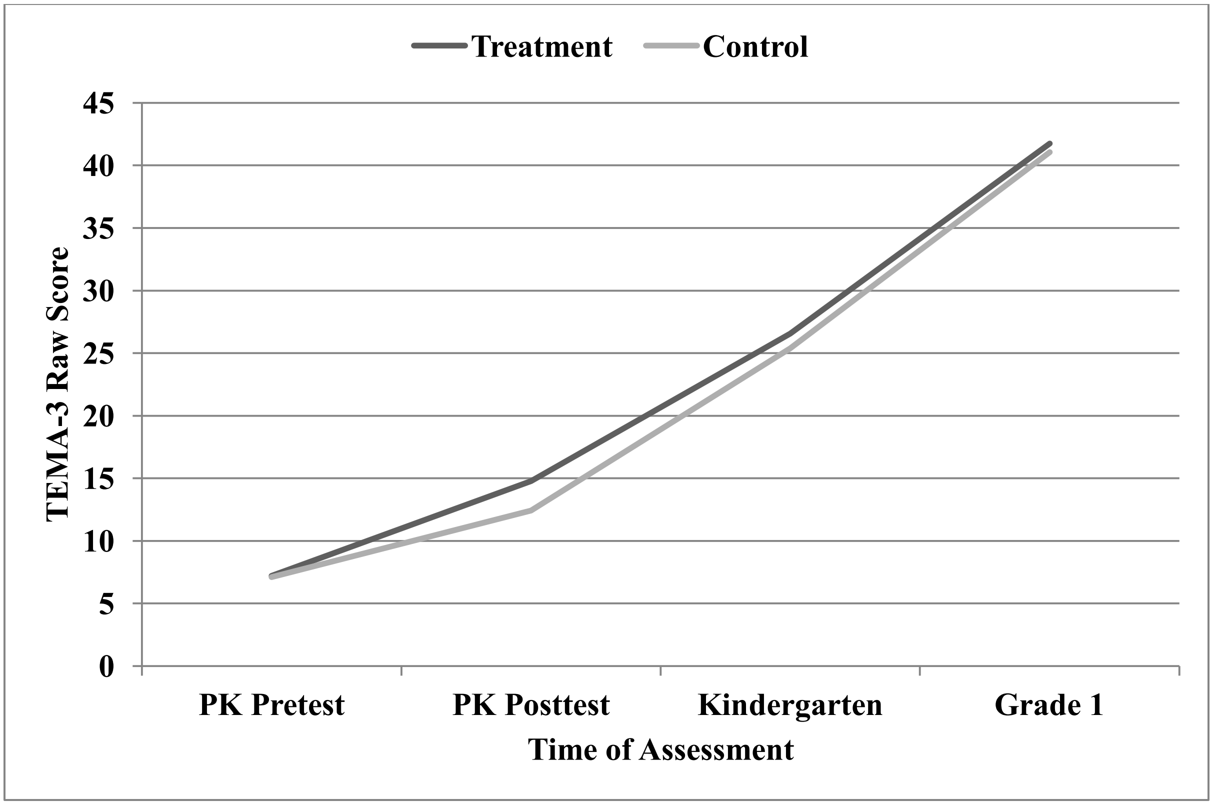

Figure 11. Main study: Growth analysis of TEMA-3 scores from pre-kindergarten entry through grade 1.

Table 13

Sustainability Study Child Demographics

	Age			Gender				Language			
				Male		Female		English		Spanish	
	n	*M*	*Range*	*n*	(%)	*n*	(%)	*n*	(%)	*n*	(%)
Treatment	326	4.37	3.82 – 5.04	150	(46.0%)	176	(54.0%)	281	(86.2%)	45	(13.8%)
CA	171	4.31	3.82 – 4.79	83	(48.5%)	88	(51.5%)	126	(73.7%)	45	(26.3%)
KY	155	4.44	3.94 – 5.04	67	(43.2%)	88	(56.8%)	155	(100.0%)	0	(0.0%)
Head Start	167	4.39	3.82 – 5.04	81	(48.5%)	86	(51.5%)	133	(79.6%)	34	(20.4%)
State Preschool	159	4.35	3.82 – 4.92	69	(43.4%)	90	(56.6%)	148	(93.1%)	11	(6.9%)
CA Head Start	95	4.32	3.82 – 4.79	47	(49.5%)	48	(50.5%)	61	(64.2%)	34	(35.8%)
KY Head Start	72	4.48	3.96 – 5.04	34	(47.2%)	38	(52.8%)	72	(100.0%)	0	(0.0%)
CA State Preschool	76	4.28	3.82 – 4.79	36	(47.4%)	40	(52.6%)	65	(85.5%)	11	(14.5%)
KY State Preschool	83	4.41	3.94 – 4.92	33	(39.8%)	50	(60.2%)	83	(100.0%)	0	(0.0%)

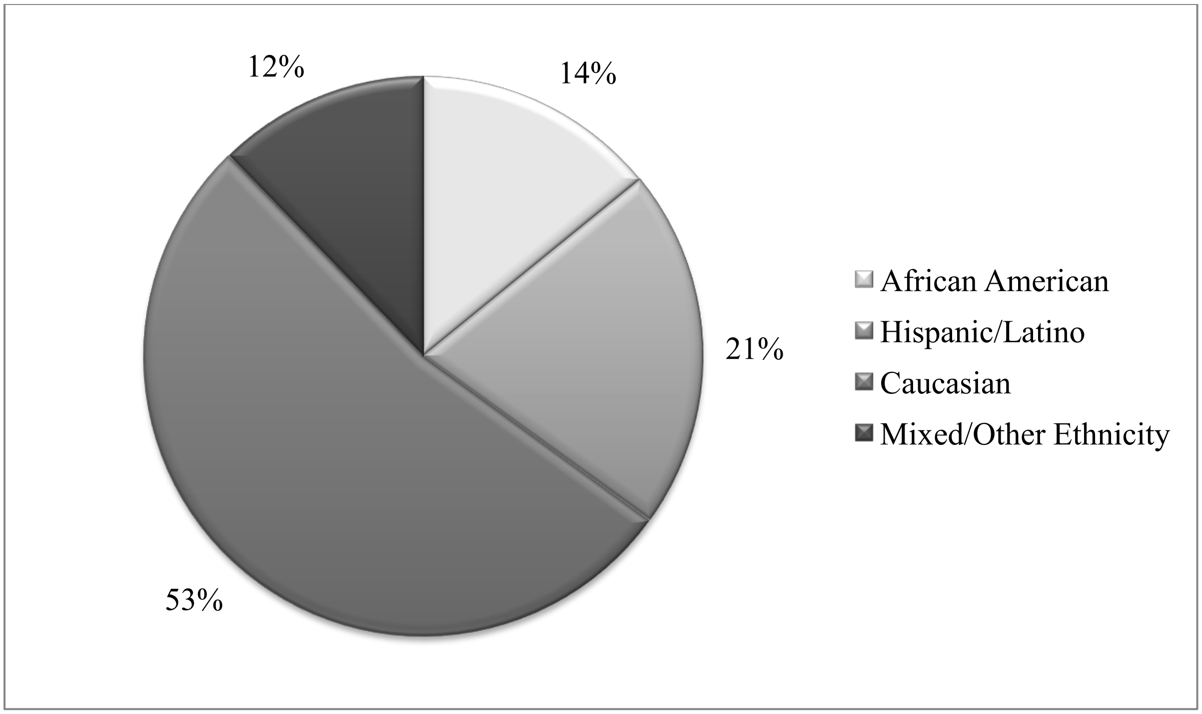

Figure 12. Sustainability Study child ethnicity.

Table 14
Sustainability Study: Fidelity of Implementation Scores

	California		Kentucky/Indiana		All Treatment
Dimension of Fidelity	State PreK	Head Start	State PreK	Head Start	Teachers
Teacher follows a schedule	0.94	0.83	0.84	0.94	.88
Teacher is prepared	0.99	0.96	0.96	0.99	.97
Teacher delivers basic activity correctly	0.97	0.94	0.94	0.98	.96
Teacher makes appropriate developmental adjustments	0.86	0.90	0.96	0.98	.92
Teacher keeps written records about each child's performance	0.92	0.86	0.93	0.95	.91
Overall fidelity	**0.94**	**0.90**	**0.93**	**0.97**	**.93**

Table 15

Implementation of the Pre-kindergarten Intervention in the Main and Sustainability Studies

	Main Study	Sustainability Study
Overall Fidelity Score	.91	.93
Mean Curriculum Dosage	44.2	43.8
Mean Home Activity Dosage	9.8	8.7
Percentage of Curriculum Activities Mastered	71	65

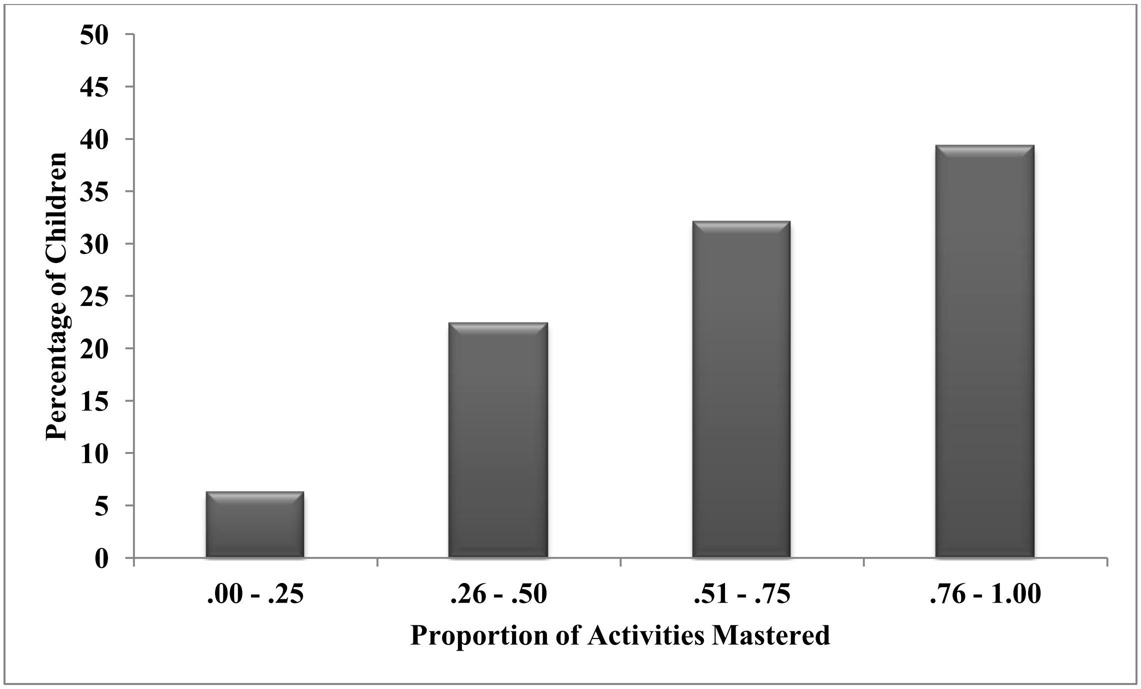

Figure 13. Percentage of children in the Sustainability Study mastering *Pre-K Mathematics* activities. Controlling for pretest scores, regression analyses revealed that mastery of the math curriculum activities significantly predicted posttest scores on both the CMA, F (2, 287) = 102.948, $p < .0001$, and the TEMA-3, F (2, 284) = 236.162, $p < .0001$.

Table 16

Sustainability Study: Mean Scores (and Standard Deviations) on the CMA and TEMA-3 in the Pre-kindergarten Year

	CMA				TEMA			
	Pretest		Posttest		Pretest		Posttest	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Treatment	.30	(.16)	.58	(.19)	7.26	(5.42)	14.78	(6.92)
CA	.29	(.17)	.55	(.20)	7.40	(5.85)	13.97	(7.00)
KY	.31	(.16)	.62	(.17)	7.11	(4.92)	15.63	(6.75)
Head Start	.29	(.16)	.55	(.19)	6.84	(4.76)	14.07	(6.06)
State PK	.31	(.17)	.62	(.18)	7.70	(6.03)	15.49	(7.63)
CA Head Start	.29	(.16)	.51	(.19)	7.13	(5.05)	12.89	(5.86)
KY Head Start	.31	(.16)	.61	(.17)	6.46	(4.34)	15.57	(6.02)
CA State PK	.30	(.18)	.61	(.19)	7.74	(6.74)	15.27	(8.00)
KY State PK	.32	(.15)	.62	(.17)	7.67	(5.33)	15.68	(7.34)
Control	.32	(.17)	.48	(.19)	7.02	(5.10)	12.49	(6.64)
CA	.31	(.16)	.47	(.17)	7.03	(5.06)	12.34	(6.03)
KY	.33	(.19)	.50	(.20)	7.01	(5.15)	12.66	(7.27)
Head Start	.32	(.17)	.49	(.19)	7.13	(5.03)	12.51	(6.70)
State PK	.31	(.18)	.48	(.18)	6.88	(5.19)	12.47	(6.58)
CA Head Start	.30	(.15)	.45	(.18)	6.68	(4.60)	11.57	(5.63)
KY Head Start	.35	(.18)	.53	(.20)	7.63	(5.45)	13.47	(7.55)
CA State PK	.32	(.17)	.50	(.17)	7.45	(5.58)	13.21	(6.38)
KY State PK	.30	(.19)	.45	(.19)	6.25	(4.69)	11.55	6.76

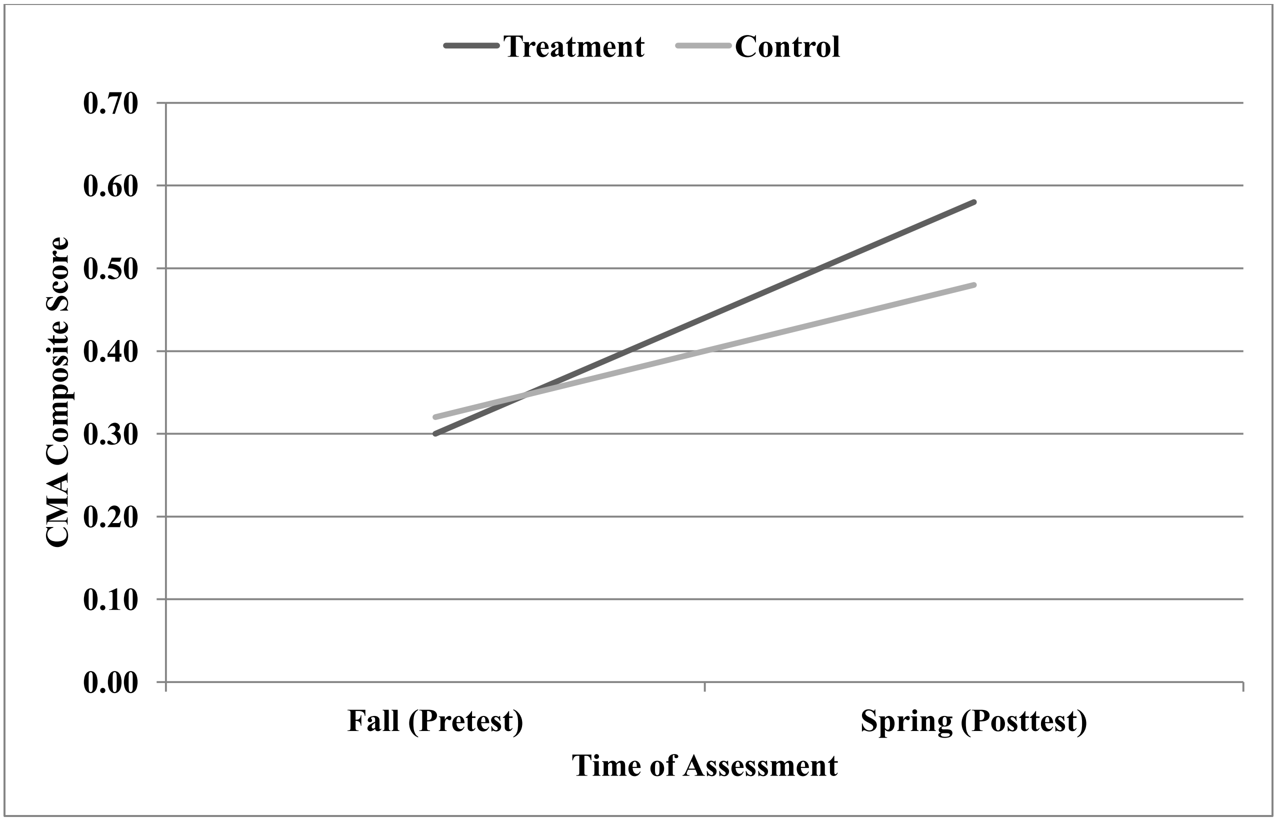

Figure 14. The effects of the Sustainability Study intervention on children's mathematical knowledge in pre-kindergarten as measured by the CMA.

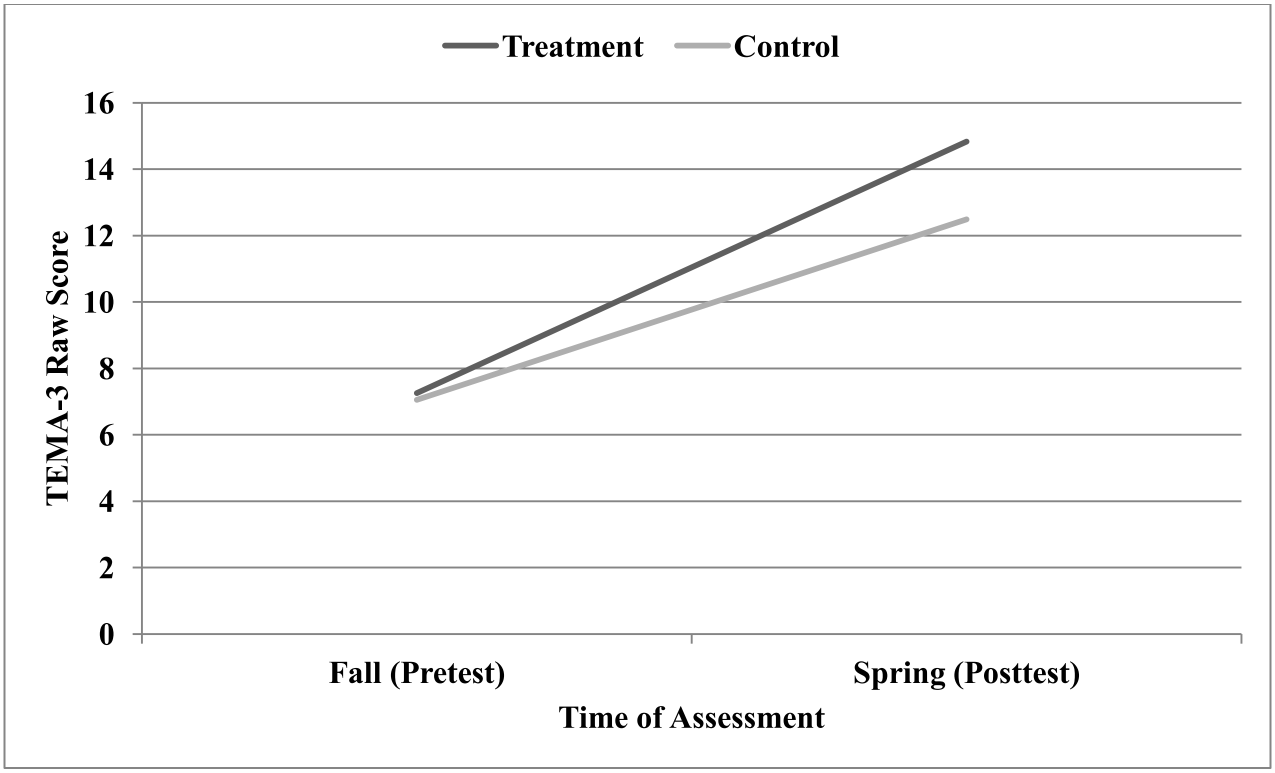

Figure 15. The effects of the Sustainability Study intervention on children's mathematical knowledge in pre-kindergarten as measured by the TEMA-3.

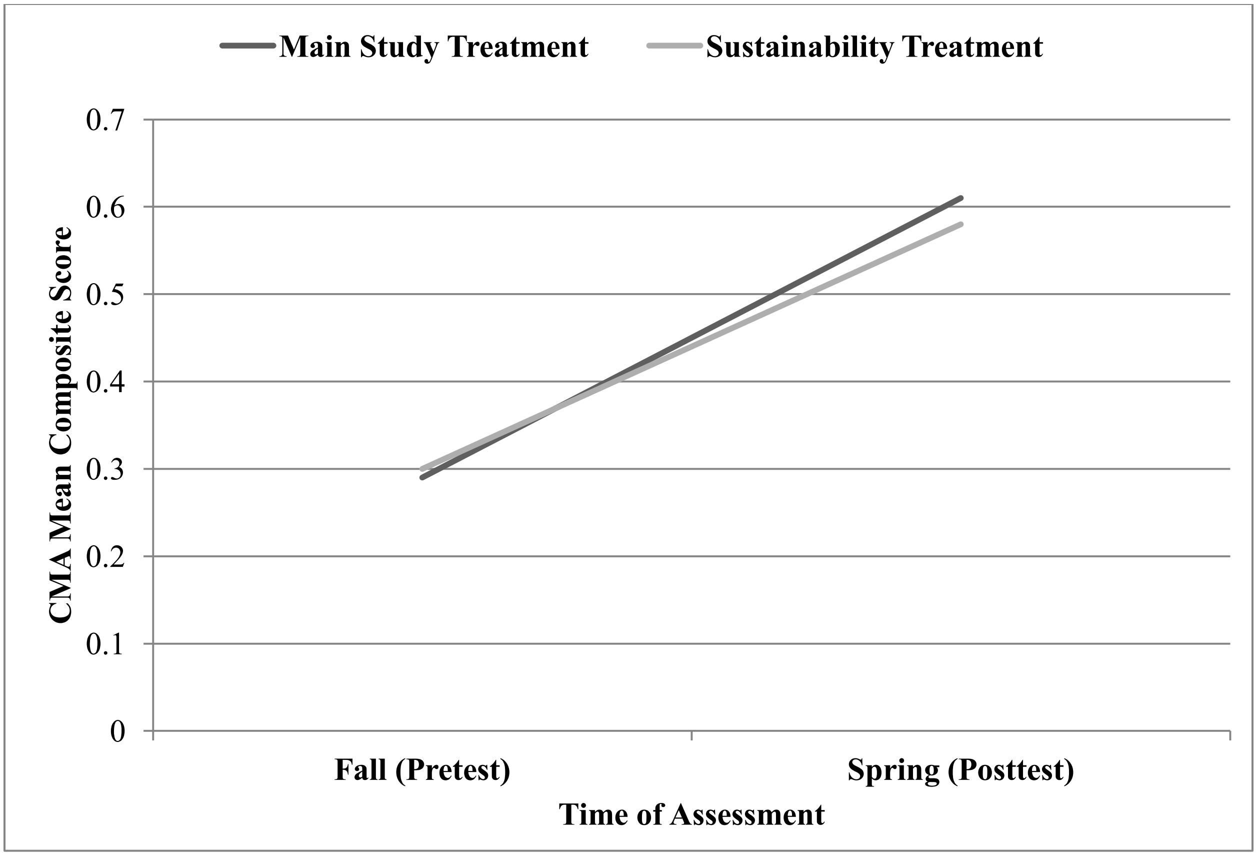

Figure 16. The effects of the main study intervention versus sustainability study intervention on children's mathematical knowledge in pre-kindergarten as measured by the CMA.

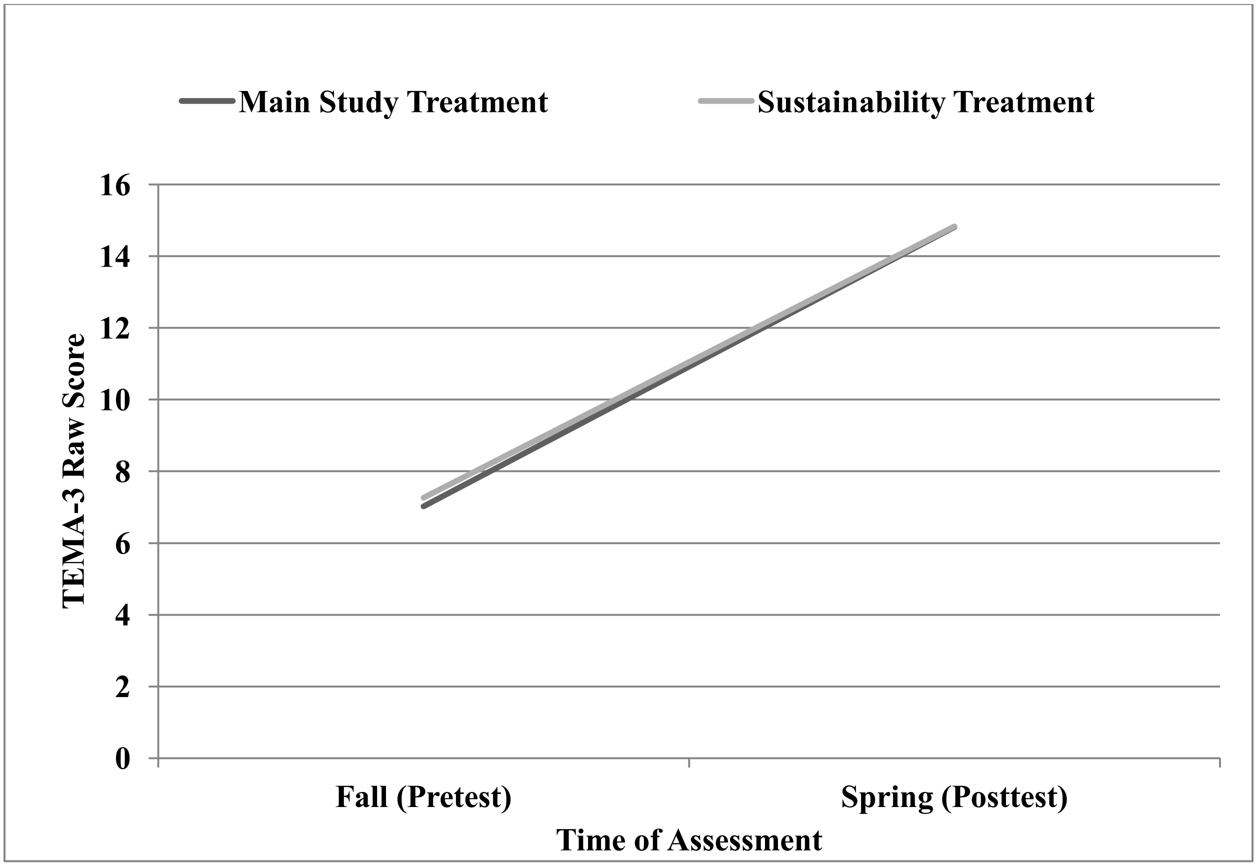

Figure 17. The effects of the main study intervention versus sustainability study intervention on children's mathematical knowledge in pre-kindergarten as measured by the TEMA-3.

Table 17
Sustainability Study: Mean Scores (and Standard Deviations) on WJ-III Understanding Directions Subtest in the Pre-kindergarten Year

	Pretest		Posttest	
	M	*(SD)*	*M*	*(SD)*
Treatment	447.23	(13.12)	456.55	(12.74)
CA	445.69	(13.15)	455.32	(13.16)
KY	448.91	(12.93)	457.80	(12.22)
Head Start	445.63	(12.33)	454.78	(12.15)
State PK	448.92	(13.74)	458.31	(13.11)
CA Head Start	443.11	(12.39)	451.98	(12.65)
KY Head Start	448.96	(11.51)	458.33	(10.56)
CA State PK	448.96	(13.42)	459.45	(12.68)
KY State PK	448.87	(14.09)	457.38	(13.47)
Control	450.37	(12.67)	455.28	(11.43)
CA	447.76	(12.81)	453.46	(10.99)
KY	453.20	(11.92)	457.27	(11.60)
Head Start	449.85	(12.37)	455.04	(11.77)
State PK	451.00	(13.03)	455.58	(11.03)
CA Head Start	446.75	(11.93)	452.20	(11.04)
KY Head Start	453.20	(12.02)	457.89	(11.86)
CA State PK	448.99	(13.77)	454.87	(10.83)
KY State PK	453.19	(11.87)	456.43	(11.29)

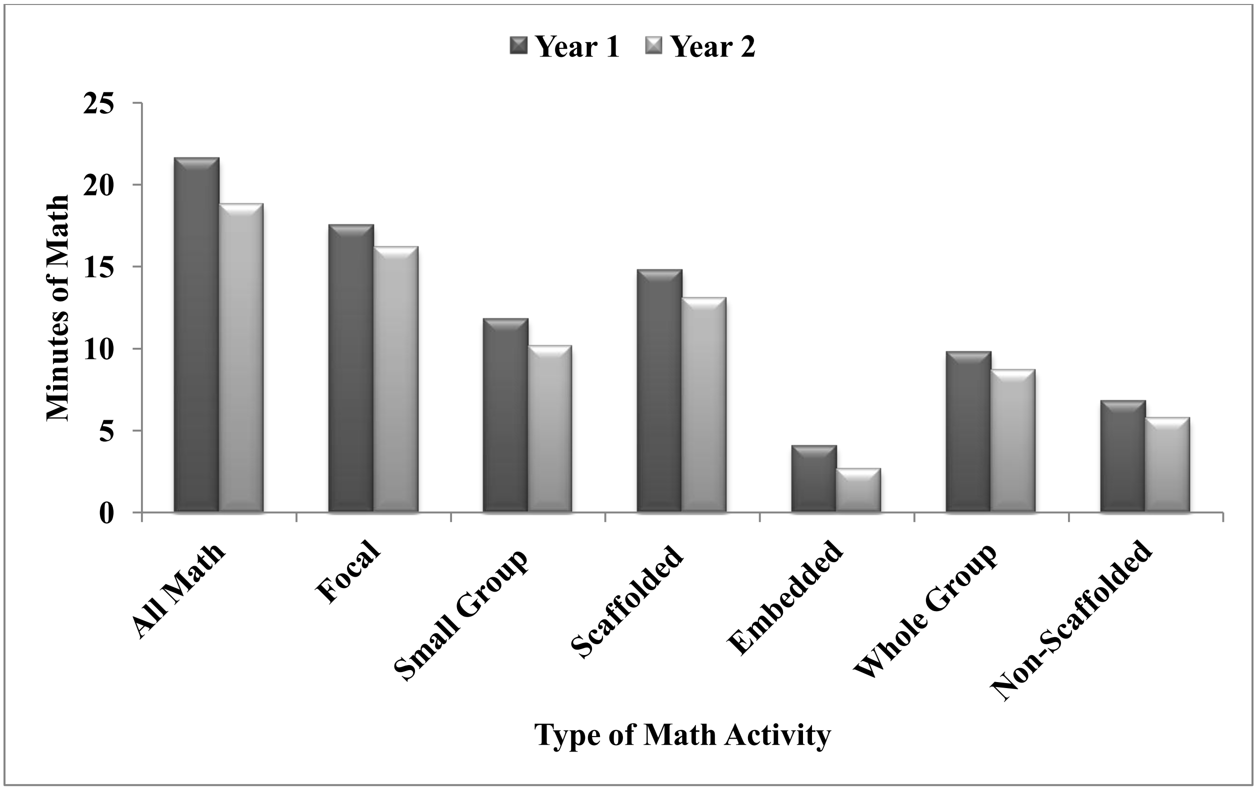

Figure 18. Math practices by pre-kindergarten treatment teachers who participated in the Main and Sustainability Studies.

Table 18

Sustainability Study Treatment Group: Mean Number of Activities Parents Reported Children Doing at Home in 1 Week

Activity	*M*	*SD*
Book and Language Activities		
Heard or read/looked at a story book	6.32	7.75
Heard or read/looked at a counting or shape book*	2.67	3.88
Heard or read/looked at a children's magazine or children's section of a newspaper	1.19	2.43
Heard or read/looked at other types of books	2.52	3.95
Heard or told children's stories	2.86	3.97
Heard or sang children's songs	6.37	7.35
Engaged in real or pretend writing activities	4.26	7.59
Store-Bought Games		
Played a board game	1.48	2.04
Played a card game	1.27	2.13
Activities Using Technology		
Watched an educational TV program or video	5.00	5.95
Played an electronic game or video game with educational content	2.46	4.54
Used a computer with children's educational software or visited a children's website with educational content	2.19	4.57
Used a computer with children's math software or visited a children's website with math content*	1.12	2.33
Educational Activities or Toys		
Played with wooden or plastic blocks	2.39	3.52
Played with puzzles, shape tiles, mazes, or other spatial materials*	2.09	2.87
Engaged in origami (paper-folding) or kitagami (paper cutting) activities*	1.37	2.24
Used materials that teach letters to children	2.49	3.55
Used materials that teach numerals to children*	2.46	3.83
Home Routine		
Helped set the table	1.96	2.85
Helped prepare food	1.47	2.31
Helped sort laundry	1.89	2.25
Activities Away from Home (Excluding Preschool)		
Attended a supplementary enrichment program for young children	1.65	2.49
Went shopping with an adult	.69	1.58
*Focal Math Activities	9.72	10.39

Table 19
Sustainability Study: Mean Scores (and Standard Deviations) on the TEMA-3 in Kindergarten

	TEMA-3	
	M	*(SD)*
Treatment	30.01	(8.33)
Control	25.67	(8.42)

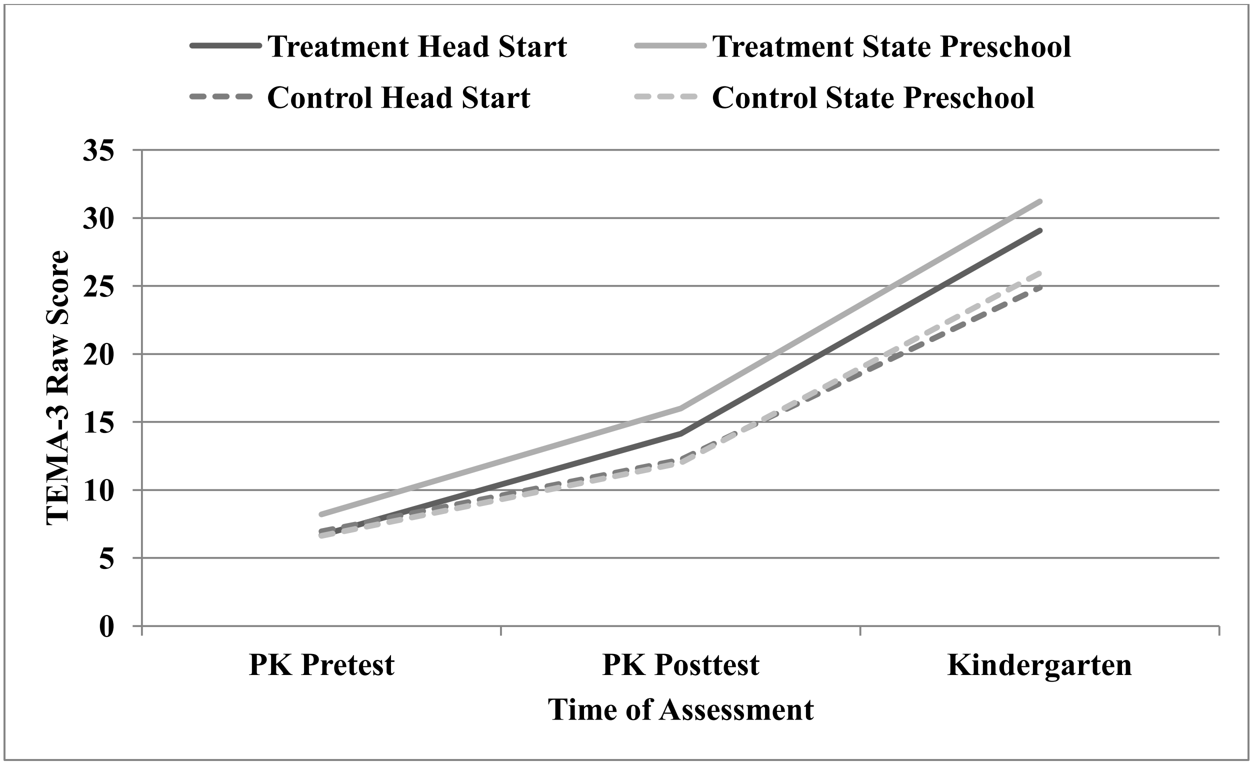

Figure 19. Sustainability Study: Growth analysis of TEMA-3 scores from pre-kindergarten entry through kindergarten by condition and type of preschool program.

Table 20
Item/Test Level Analysis of the Child Math Assessment

Based on 39 Items

	Proportion Correct (in range)	Average Proportion Correct	KR-20
Pretest	0.02 - 0.66	0.31	0.87
Posttest	0.16 - 0.94	0.55	0.89

Based on 10 Tasks

	Stratified Alpha
Pretest	0.91
Posttest	0.92

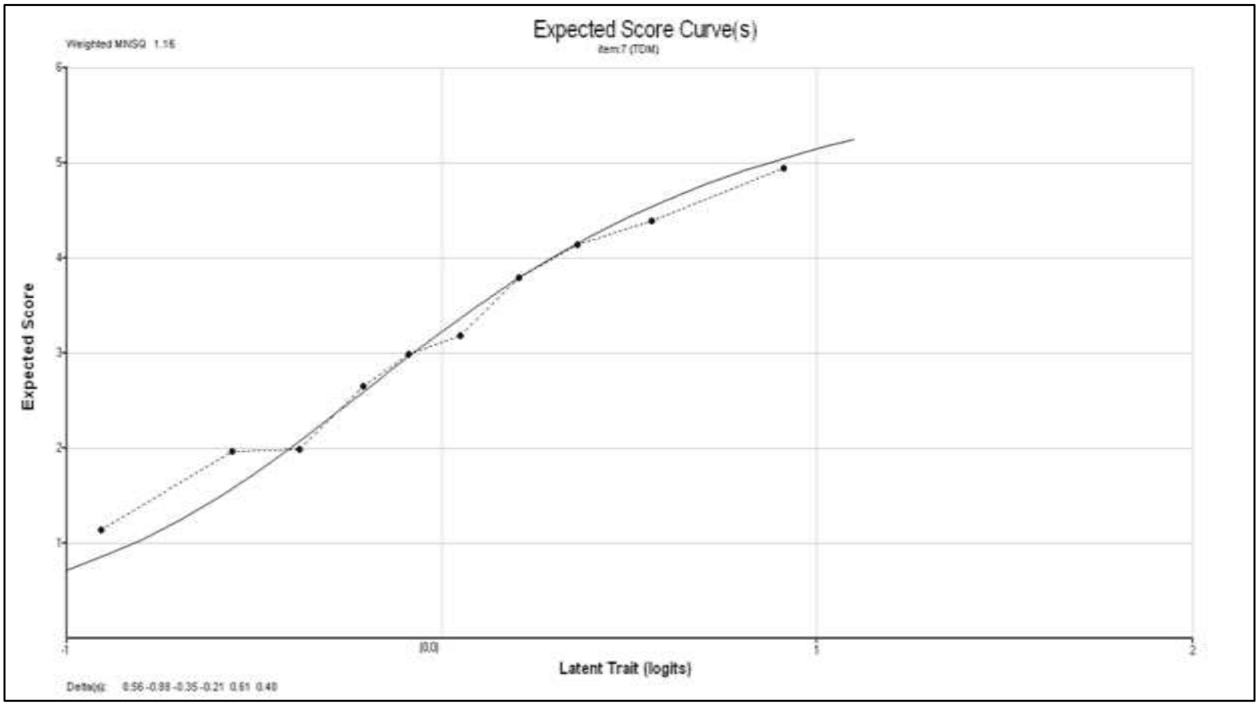

Figure 20. Expected score curve for direct measurement.