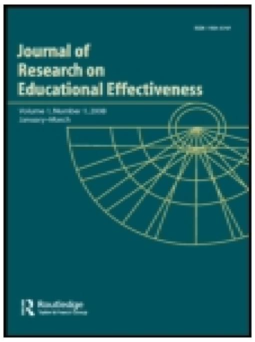


Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

Longitudinal Evaluation of a Scale-Up Model for Teaching Mathematics With Trajectories and Technologies

Julie Sarama , Douglas H. Clements , Christopher B. Wolfe & Mary Elaine Spitler

Journal of Research on Educational Effectiveness, 5: 105–135, 2012

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345747.2011.627980

INTERVENTION, EVALUATION, AND POLICY STUDIES

Longitudinal Evaluation of a Scale-Up Model for Teaching Mathematics With Trajectories and Technologies

Julie Sarama, Douglas H. Clements, Christopher B. Wolfe, and Mary Elaine Spitler
University at Buffalo, SUNY, Buffalo, New York, USA

Abstract: We used a cluster randomized trial to evaluate the effectiveness of a research-based model for scaling up educational interventions, focusing on the persistence of effects with and without a follow-through intervention. The instantiation of the Technology-enhanced, Research-based, Instruction, Assessment, and professional Development (TRIAD) model emphasized teaching early mathematics for understanding via learning trajectories and technology. The TRIAD implementation began in 42 schools in two city districts serving low-resource communities, randomly assigned to three conditions. In pre-kindergarten, the 2 experimental interventions were identical, but 1 included follow-through in the kindergarten year, including knowledge of the pre-K intervention and ways to build upon that knowledge using learning trajectories. Intent-to-treat analyses showed that students in both the follow-through condition ($g = .33$) and non-follow-through condition ($g = .22$) scored statistically significantly higher than children in the control condition. Both groups outperformed the control condition in treatment-on-the-treated analyses ($g = .38$, follow-through; $g = .30$ non-follow-through). Moderators and mediators were also analyzed. We conclude that the instantiation of the TRIAD model was successful and that follow through may contribute to the persistence of the effects of preschool interventions.

Keywords: Mathematics, scale up, early childhood, learning trajectories, longitudinal

Education needs transferable, practical examples of scale up (McDonald, Keesler, Kauffman, & Schneider, 2006), empirical evidence of the effectiveness of these examples, and focused research on critical variables—all leading to refined, generalizable theories and models of scale up. A challenging practical and theoretical issue is scaling up successful mathematics interventions in the early and primary grades in the United States and especially ensuring the persistence of the effects of such interventions. We created a research-based model to meet this challenge in the area of early mathematics, with the intent to generalize the model to other subject matter domains and other grade levels. This article presents the results of an experimental evaluation of the 1st year of implementation of a follow-through intervention.

The specific implementation of the TRIAD (*T*echnology-enhanced, *R*esearch-based, *I*nstruction, *A*ssessment, and professional Development) model (Sarama, Clements,

Address correspondence to Douglas H. Clements, Learning and Instruction, Graduate School of Education, 212 Baldy Hall (North Campus), University at Buffalo, SUNY, Buffalo, NY 14260, USA. E-mail: clements@buffalo.edu

Starkey, Klein, & Wakeley, 2008) was designed to increase mathematics achievement in young children, especially those at risk. All aspects of the intervention—mathematical content, pedagogy, technology, and assessments—are based on a common core of learning trajectories. The rationale for the creation of the TRIAD model was the confluence of educational needs in mathematics and early childhood. U.S. proficiency in mathematics is low (Kilpatrick, Swafford, & Findell, 2001; Mullis et al., 2000; National Mathematics Advisory Panel, 2008). From a human capital perspective, the most is gained from interventions with the youngest children (Carneiro & Heckman, 2003). Moreover, students who live in poverty and are members of linguistic and ethnic minority groups have particularly low levels of achievement (Denton & West, 2002; National Mathematics Advisory Panel, 2008; National Research Council, 2001, 2009), because they receive less support in their home and school environments (Blevins-Knabe & Musun-Miller, 1996; Ginsburg & Russell, 1981; Griffin, Case, & Capodilupo, 1995; Holloway, Rambaud, Fuller, & Eggers-Pierola, 1995; Jordan, Huttenlocher, & Levine, 1992; Paris, Morrison, & Miller, 2006; Sarama & Clements, 2009). For this same reason, high-quality experiences appear especially beneficial for children living within low-resource communities (Carneiro & Heckman, 2003; Raudenbush, 2009).

The TRIAD implementation began in 42 schools in two city districts serving low-resource communities. Schools were randomly assigned to one of three conditions. At the preschool level (called pre-K hereafter to denote the pre-kindergarten year), the two experimental conditions were identical. Evaluations revealed a substantial and significant effect at the end of pre-K (Clements, Sarama, Spitler, Lange, & Wolfe, 2011). One experimental condition was assigned to experience a follow-through intervention into the kindergarten year, in which teachers were taught about the pre-K intervention and ways to build upon it using learning trajectories. This article evaluates the persistence of the effects of this implementation, with and without TRIAD's follow-through intervention in the kindergarten year. We also examined whether these effects were equivalent for various subpopulations, as well as possible indirect or mediational effects through particular pedagogical practices, such as those identified for the pre-K intervention (Clements et al., 2011).

THEORETICAL FRAMEWORK AND TRIAD'S RESEARCH-BASED GUIDELINES FOR SCALING UP

TRIAD's theoretical framework (Sarama et al., 2008) is an elaboration of the *Network of Influences* framework (Sarama, Clements, & Henry, 1998). Scale up is considered to involve multiple coordinated efforts to maintain the integrity of the vision and practices of an innovation through increasingly numerous and complex socially-mediated filters. The depiction of the Network of Influences framework in Figure 1 illustrates the hypothesized influences of context and implementation variables on outcomes, such as teacher knowledge and child achievement. Briefly, communication, collaboration, and agreement among all groups are essential (Figure 1, oval I; Sarama et al., 1998). Research also suggests the importance of a knowledgeable and responsive teacher and thus of intense professional development (Figure 1, J, N; Darling-Hammond, 1997; National Research Council, 2009; Sarama & DiBiase, 2004), particularly that is focused on research-based models of children's thinking and learning (Bredekamp, 2004; Carpenter & Franke, 2004; Hiebert, 1999; Klingner, Ahwee, Pilonieta, & Menendez, 2003; Lawless & Pellegrino, 2007). Efficacious teacher-training strategies include demonstrations, practice, and feedback, especially from

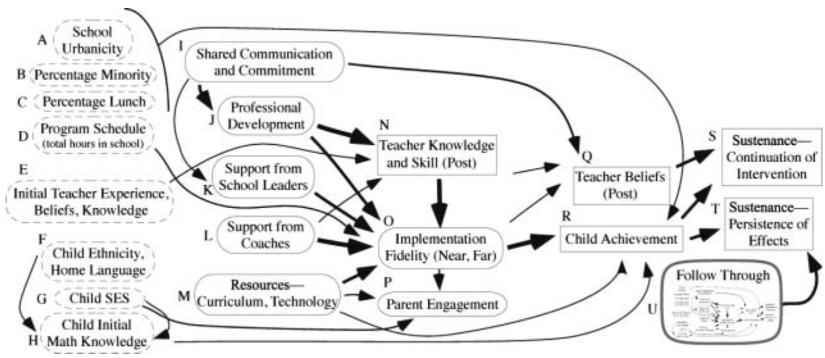


Figure 1. Revised Network of Influences theoretical framework. *Note.* The "Follow Though" model at the bottom right, most relevant to this study, is simple a microcosm of the Framework. *Contextual variables* in dotted ovals include the school (A-D), teacher (E), and child (F-H) factors. For example, child socioeconomic status, or SES (G), impacts children's initial mathematics knowledge (H), which influences children's achievement (R)—an outcome variable indicated by the solid rectangle. *Implementation variables* in solid ovals are features that the project can encourage and support, but cannot control absolutely. For example, **heavy arrows** from professional development (J), to teacher knowledge (N), to implementation fidelity (O), to child achievement (R), indicate the strong effects in that path. Support from coaches (L) also has a strong effect on implementation fidelity, while other factors (J, K, M) are influential, but to a moderate degree (not all small effects are depicted). Relationships are further described in the following section.

coaches (e.g., strong effects from J and L to N and O in Figure 1; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; Pellegrino, 2007; Showers, Joyce, & Bennett, 1987).

Building on this theoretical and empirical research base, we established relevant guidelines that define the TRIAD model, which we briefly summarize here (see Sarama et al., 2008, for the full model, including research citations and the 10 guidelines). The model involves communication among key groups, especially school leaders at all levels, around a shared vision of the innovation (Figure 1, I–M; Berends, Kirby, Naftel, & McKelvey, 2001; Borman, Hewes, Overman, & Brown, 2003; Elmore, 1996; Glennan, Bodilly, Galegher, & Kerr, 2004; Hall & Hord, 2001), with frequent formative assessment ("checking up") that involves the key groups in cycles of data collection and problem solving (Figure 1, I, Q, R, S, T; Fixsen, Blase, Naoom, & Wallace, 2009; Hall & Hord, 2001; Huberman, 1992; Kaser, Bourexis, Loucks-Horsley, & Raizen, 1999; Snipes, Doolittle, & Herlihy, 2002). The professional development is ongoing, intentional, reflective, situated in the classroom and the school, grounded in particular curriculum materials, and focused on mathematics content knowledge and children's thinking, especially on learning trajectories (Figure 1, J–M; Bodilly, 1998; Borman et al., 2003; Clements & Sarama, 2004; Cohen, 1996; Elmore, 1996; Guskey, 2000; Hall & Hord, 2001; Kaser et al., 1999; National Mathematics Advisory Panel, 2008; Raudenbush, 2008). We defined learning trajectories as "descriptions of children's thinking and learning ... and a related, conjectured route through a set of instructional tasks" (Clements & Sarama, 2004, p. 83). Thus, learning trajectories have three components: a goal (i.e., an aspect of a mathematical domain children should learn); a developmental progression, or learning path through which children move through levels of thinking; and instruction that helps them move along that path. Learning trajectories facilitate teachers' learning about mathematics, how children think about and learn mathematics, and how such learning is supported by the curriculum and its teaching strategies.

They address domain-specific components of learning and teaching that have the strongest impact on cognitive outcomes (Lawless & Pellegrino, 2007). By illuminating potential developmental progressions, they bring coherence and consistency to goals, curricula, and assessments (for a comprehensive description and review of research on these learning trajectories, see Sarama & Clements, 2009, and for the instructional component, see Clements & Sarama, 2009).

The present project, the largest-scale implementation of the TRIAD model to date, was found to be successful at the pre-K level on near and far transfer measures. Children in the TRIAD groups learned more mathematics than the control children (effect size, $g = 0.72$, Clements et al., 2011). Of possible moderators (including school percentage of free/reduced lunch and limited English proficiency, and child level gender and Individual Education Plan status variables), only one was significant: African American children learned less than other children in control classrooms and more than other children in TRIAD classrooms. Three components of a measure of the quantity and quality of classroom mathematics environments and teaching partially mediated the treatment effect: total number of computers on and working for children, the classroom culture component, and the total number of math activities (Clements et al., 2011). The pre-K TRIAD implementation also showed far transfer effects on measures of language competence in the beginning of their kindergarten year (Sarama, Clements, Lange, & Wolfe, in press).

LONGITUDINAL EFFECTS, THE ISSUE OF "FADING," AND THE NEED FOR FOLLOW THROUGH

Longitudinal evaluation is particularly important for the pre-K and primary years. Lasting effectiveness can be categorized as *sustainability* or *persistence.* We use sustainability to mean the length of time an innovation continues to be implemented with fidelity (cf. Baker, 2007), the topic of a different TRIAD study. We use persistence to mean the continuation of the effects of an intervention on individual children's trajectories of learning, the focus of this study.

Some studies indicate that early interventions can have persistent effects. For example, several have shown positive and long-lasting effects of preschool experience (Broberg, Wessels, Lamb, & Hwang, 1997; Gray, Ramsey, & Klaus, 1983; Magnuson & Waldfogel, 2005; Montie, Xiang, & Schweinhart, 2006). However, there is considerable empirical research and resultant (practical) assertions that preschool gains fade in the primary grades. For example, in one study of six cohorts, gains in preschool weakened as children progressed through the primary grades, disappearing by fourth grade (Fish, 2003). Other studies show a similar fade (Natriello, McDill, & Pallas, 1990; Preschool Curriculum Evaluation Research Consortium, 2008; Turner & Ritter, 2004; U.S. Department of Health and Human Services—Administration for Children and Families, 2010).

Although an ostensible reason is that early effects are themselves evanescent, we believe that a contradictory explanation is more theoretically cogent. That is, we hypothesize that most present educational contexts (e.g., curricula, minimal requirements, teaching practices) are unintentionally aligned against the persistence of early interventions. Consider the educational trajectories of children who benefited from a successful pre-K experience as they move into kindergarten. The kindergarten curriculum they experience likely assumes little or no mathematical competence, so only low-level skills are taught. Their teachers are often required to follow such curricula rigidly and remain unaware that some of their students have already mastered the material they are about to "teach" (Bennett, Desforges, Cockburn, & Wilkinson, 1984; Clements & Sarama, 2009; National Research Council, 2009; Sarama & Clements, 2009). Further, biases may negatively affect the subsequent school experiences of children at risk during pre-K. For example, kindergarten teachers rated Head Start children's mathematics ability as lower than that of other children, even though direct assessments showed no such differences (U.S. Department of Health and Human Services—Administration for Children and Families, 2010). Even if the children are assigned to a kindergarten teacher who recognizes their competencies, pressure to increase the number of children passing minimal competency assessments may lead this teacher to work mainly with (and/or mainly at the level of) the lowest performing children. Within this context and without continual, progressive support, early gains are frequently "lost." In this way, we believe the present U.S. educational system unintentionally but insidiously reopens the gap between students from low- and higher resource communities.

For these reasons, we designed and evaluated the effectiveness of TRIAD's follow-through intervention, testing our hypothesis that such follow through is the "missing piece" in many early interventions whose longitudinal evaluations have found less positive effects. We agree that

> it is unrealistic, given our knowledge of development, to expect short-term early interventions to last indefinitely, especially if children end up attending poor quality schools. It is magical thinking to expect that if we intervene in the early years, children will need no further help in the elementary school years and beyond. (Brooks-Gunn, 2003, p. 1)

Although this might appear to be an issue of simple "educational engineering," the issue has momentous implications for both theory and policy. Interpretations of this "fade" often call for *decreased* funding and attention to pre-K (Fish, 2003, 2007). Although this may appear reasonable (with logic such as, if effects fade out, why fund that intervention?), we believe this mistakenly treats initial effects of interventions as independent of the future school contexts. Instead, we believe children's trajectories must be studied as they experience different educational courses. If such effects "fade" in traditional settings but do not in the context of follow-through interventions, then attention to and funding for follow-through efforts for both pre-K and the primary grades should arguably increase.

Kindergarten teachers in schools assigned to TRIAD's follow-through condition were taught about the pre-K intervention and ways to build upon it. That is, they were shown the mathematics many of their entering students had learned. Teachers were also taught about the learning trajectories extended through the kindergarten curriculum, including the developmental progressions and how to modify their extant curricula to more closely match the levels of thinking of their students. They also received access to the *Building Blocks* software (Clements & Sarama, 2007a), which follows the learning trajectories through the primary grades and is the same suite that the students had used in pre-K.

RESEARCH QUESTIONS

There were three primary research questions.

1. *What is the persistence of effects of the TRIAD intervention, with and without follow-through, on achievement at the end of kindergarten?* Do children in the TRIAD

follow-through (TRIAD-FT) group on the average outperform children in the TRIAD non-follow-through (TRIAD-NFT) group in mathematics achievement at the end of kindergarten? Do children in each of these experimental groups outperform those in the control group?

2. *Are there significant moderators of any statistically significant effects?* Do the effects of the three conditions differ by gender, disability status, or ethnic group?
3. *Do measures of teachers' practice mediate effects of different treatments on achievement?* If the effects are significant, do specific measures of the quantity and quality of the kindergarten mathematics environment and teaching mediate the effects on mathematics achievement?

METHOD

We used a cluster randomized (at the school level) experimental design that enabled a formal test of the generalizability of TRIAD's impact over the varied settings in which it was implemented. We made a list of eligible schools (all those who had not participated in prior *Building Blocks* or TRIAD research or development projects) and the statewide mathematics achievement scores of fourth graders who attended those schools in the year prior to randomization. Based on these fourth-grade math scores, schools were ranked from highest to lowest scores separately within each site. Blocks were created to contain three schools with similar fourth-grade math scores in descending order. We (publically, with five observers, including school administrators and project staff members) then assigned each eligible school to one of three treatment groups, selected randomly from the blocks, using a table of random numbers. We employed hierarchical linear models (HLMs) to calculate the effects of the interventions on students' mathematics performance and to account for possible variations of the effects among groups. Such a procedure is most appropriate for studies of scale up, especially within longitudinal designs (Raudenbush, 2007, 2008).

Participants and Contexts

Participants were the 1,305 students from the original 42 schools in Buffalo, New York, and Boston, Massachusetts, who had both a pretest and posttest in pre-K (Clements et al., 2011), and the kindergarten teachers in those schools. Table 1 describes these diverse populations. By the end of kindergarten, 1,218 of these children were available for both components of the mathematics achievement test. This is the population we used for intent-to-treat (ITT) analyses in which the condition group (school) to which the child was originally assigned was maintained as that child's condition, regardless of how many days the child experienced in that school. These analyses estimate potential effects of the intervention from a policy perspective. Of these children, 963 remained within their randomized school. We used this latter population for treatment-on-the-treated (TOT) analyses that estimate the effects for children who experienced the full duration of the condition to which they were originally assigned. Of the 21% of the 1,218 students who were not included in the TOT, 37% were originally assigned to the TRIAD-FT research group, 33% were assigned to the TRIAD-NFT group, and 30% were assigned to the control group. Of the students assigned to the intervention groups, 84% (FT) and 81% (NFT) transferred to nonstudy schools either immediately after the pre-K year or during the K year. Thus, they had varying

Table 1. Demographics of participating schools

Name	No. of Schools	*M* No. of Children per School	*M* No. of Children per Classroom	No. of Children	% Female	NA	AA	A/P	H	W	Other	% Free/ Reduced Lunch	English Language Learning
All	42	31.07	19.39	1305	51%	24 (1.8%)	695 (53.3%)	48 (3.7%)	282 (21.6%)	248 (19.0%)	8 (.60%)	82.33	13.46
TRIAD-FT	12	39.25	18.91	471	52.70%	13 (2.8%)	275 (58.4%)	20 (4.2%)	92 (19.5%)	69 (14.6%)	2 (.40%)	81	13.28
TRIAD-NFT	14	32.57	20.23	456	49.80%	4 (.90%)	238 (52.2%)	9 (2.0%)	90 (19.7%)	112 (24.6%)	3 (.70%)	83.57	10.5
Control	16	23.63	19.17	378	50.30%	7 (1.9%)	182 (48.1%)	19 (5.0%)	100 (26.5%)	67 (17.7%)	3 (.80%)	82.25	16.18

Note. In some cases, records allowed no further categorical breakdowns. NA = Native American; AA = African American; A/P = Asian/Pacific Islander; H = Hispanic; W = White non-Hispanic; TRIAD-FT = TRIAD intervention in pre-K and follow through in kindergarten; TRIAD-NFT = TRIAD intervention in pre-K only.

amounts of exposure to the intervention after the pre-K year. The remaining intervention students transferred to schools in other study conditions, including the control condition, or skipped kindergarten. A one-way analysis of variance revealed that the ITT and TOT groups did not differ significantly on kindergarten outcome scores, $F(1, 1217) = .001$, $p = .973$.

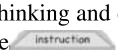
The year that these children were in kindergarten, one district adopted a substantially revised version of their mathematics curriculum, the kindergarten level of Investigations (*Investigations,* 2008), whereas the other district continued to use the first edition of the same curriculum. Both districts wrote and disseminated "pacing guides" that established what unit of the curriculum should be taught each week. For example, specific lessons were to be taught on specific days (e.g., lessons from the curriculum's "Day 1" on October 7, lessons from "Day 2" on October 8 . . .). The "walk through" form used by administrators included items on this pacing guide.

The TRIAD-FT Intervention

TRIAD staff worked with 43 TRIAD-FT kindergarten teachers for approximately seven sessions (a maximum of approximately 32 hr), spread over the intervention year. We computed the percentage of participation in professional development based on the percentage of hours attended out of the maximum number of hours offered to any participant during that program year. Teacher attendance ranged from 16% to 100% (M = 63%, Mdn = 63%). School-level attendance ranged from 31% to 90% (M = 63%, Mdn = 60%). Professional development was conducted at the university and at district/community buildings designated for professional development. Prior to the first professional development session, project staff met with kindergarten and pre-K teachers on site at each follow-through school to facilitate an exchange of information between the experimental pre-K teachers and the kindergarten teachers regarding the particular mathematics knowledge and skills of students who had participated in the *Building Blocks* curriculum during the preceding year.

We attempted to follow research on training that is both effective and attractive to teachers. Teachers prefer training that meets at least once per month, provides curriculum materials, and engenders a sense of "personal satisfaction" (Sarama & Clements, 2009, pp. 348–349; Sarama & DiBiase, 2004). They prefer pedagogical strategies of demonstration/modeling and small-group discussion, followed by handouts, lecture, observing actual practice, games, role-play, and video (Wolfe, 1991).

Because the kindergarten teachers already had a curriculum, the focus on teaching kindergarten mathematics was not on the introduction of a new curriculum (as it had been in pre-K) but rather on the empirically supported strategy of formative assessment (e.g., Foorman, Santi, & Berger, 2007; National Mathematics Advisory Panel, 2008). Teachers first studied the content and developmental progressions for each major mathematical topic. Approaches to this included presentations and discussions, as well as the use of the software application, *Building Blocks Learning Trajectories* (BBLT), which presented and connected the components of the innovation. BBLT provided scalable access to the learning trajectories via descriptions, videos, and commentaries. Two sequential aspects of the learning trajectories—the developmental progressions of children's thinking and connected instruction—are linked to the others (see Figure 2). As an example, using these approaches, teachers studied the learning trajectory for counting. Initial presentations including viewing the levels of thinking that constitute the developmental progression, using the BBLT and

Figure 2. Building Blocks Learning Trajectories (BBLT) web application. BBLT provides scalable access to the learning trajectories via descriptions, videos, and commentaries. Each aspect of the learning trajectories—*developmental progressions* of children's thinking and connected *instruction*—are linked to the other. For example, teachers might choose the  (curriculum) view and see the screen on the left, below. Clicking on a specific activity provides a description. Clicking on  slides the screen over to reveal descriptions, several videos of the activity "in action," notes on the video, and the level of thinking in the learning trajectory that activity is designed to develop, as shown below on the right. (See UBTRIAD.org for a demonstration.) Clicking on the related developmental level, or child's level of thinking, ringed above, switches to the view *of that topic and that level of thinking*. This likewise provides a description, video, and commentary on the developmental level—the video here is of a clinical interview task in which a child displays that *level of thinking*. Teachers can also study a development view, studying clinical interviews of children at each level of thinking, and, if desired, link back to activities.

discussing the attributes, both mathematical and psychological, of each level. As an illustration, these included the cardinal principle (the last counting word "tells how many"), which is often underemphasized in curricula and teaching, and, at higher levels, competencies in counting on and other counting strategies. Teachers used the "Test Yourself" feature of the BBLT to evaluate their abilities to diagnose children's level of counting (by identifying the level evinced by children in randomly selected videos). They also used the BBLT's links to view research-based instructional strategies to promote children's progress to the next level. Teachers worked in small groups to plan how activities from their curriculum might promote, or be modified to promote, learning for the relevant levels. They also planned how to integrate the *Building Blocks* software, the only published curricular component included in the TRIAD-FT intervention, into their existing curriculum. Thus, the TRIAD-FT intervention was instantiated within the teachers' prescribed curriculum but was presented in a manner that could be used with any curriculum.

Measures

School Environment and Teaching in Mathematics. An observational instrument measured the quality and quantity of the classrooms' environments and teaching practices. The Classroom Observation of Early Mathematics Environment and Teaching (COEMET) was created based on a body of research on the characteristics and teaching strategies of effective teachers of early childhood mathematics and has been employed in previous research (Clements & Sarama, 2008b; Clements et al., 2011; Sarama et al., 2008). There are 28 items, all but four of which are 5-point Likert scales. The COEMET was designed to document how mathematics is taught and what happens in each classroom (Hall & Hord, 2001).

Assessors spend no less than a half day in the classroom, for example, from before the children arrive until the end of the half day (in this study, about half were conducted after lunch to the end of the school day, if that were the period during which mathematics was taught). All mathematics activities are observed and evaluated. The COEMET subscales include classroom culture, specific mathematics activities (SMAs), the overall number of specific math activities observed, and the total time during the observation devoted to mathematics instruction. Assessors complete the classroom culture section once to reflect their entire observation. They complete an SMA form for each observed mathematics activity, defined as one conducted intentionally by the teacher involving several interactions with one or more children, or set up or conducted intentionally to develop mathematics knowledge.

Interrater reliability for the COEMET, computed via simultaneous classroom visits by pairs of observers (10% of all observations, with pair memberships rotated) was 88%; 99% of the disagreements were the same polarity (i.e., if one was *agree*, the other was *strongly agree*). Coefficient alpha (interitem correlations) ranged from .95 to .97 for the total instrument, with .74 for the Classroom Culture scale and .90 for the SMA scale (Clements & Sarama, 2008a; Clements et al., 2011). In prior research, predictive validity was supported by a regression in which the Classroom Observation total score accounted for a significant amount of the variance in children's posttest achievement scores ($t = 2.52$, $p < .05$) after accounting for pretest scores (Clements & Sarama, 2008a).

Children's Mathematical Knowledge. The Research-based Elementary Math Assessment (REMA; Clements, Sarama & Liu, 2008) measures core mathematical abilities of children from age 3 to 8 years using an individual interview format, with standardized protocol and scoring procedures. Abilities are assessed according to theoretically and empirically based developmental progressions that underlie the *Building Blocks* learning trajectories. Developmental progressions in number include verbal counting, object counting, subitizing, number comparison, number sequencing, connection of numerals to quantities, number composition and decomposition, adding and subtracting, and place value. These items help distinguish between children who have not constructed true number concepts and those who have. Geometry progressions include shape recognition, shape composition and decomposition, congruence, construction of shapes, and spatial imagery, as well as geometric measurement and patterning. The REMA defines mathematical competence as a latent trait in item response theory, yielding a score that locates children on a common ability scale with a consistent, justifiable metric (allowing accurate comparisons, even across ages and meaningful comparison of change scores, even when initial scores differ; B. D. Wright & Stone, 1979). All items are ordered by Rasch item difficulty; children stop the assessment after four consecutive errors. Explicit protocols and procedures exist for administration, videotaping, coding, scoring, and staff training on all aspects. All sessions were videotaped, and each item coded for accuracy and, when relevant, for solution strategy by two trained coders. Any discrepancies were resolved via consultation with the senior researchers. Continuous coder calibration militated against drift. Previous analysis of the assessment data showed that its reliability ranged from .75 to .94 on the subtests and .93 to .94 on the total test scores (see Clements et al., 2008, for full details on content and concurrent validity); on the present population, the reliability was .92. Support for the instrument's validity was provided by Rasch analyses (Clements et al., 2008), with good fit indicating construct validity (Bond & Fox, 2001). In addition, concurrent validity was established with a .86 correlation with a different measure of preschool mathematics achievement (Clements et al., 2008) and a correlation of .74 with the Woodcock–Johnson Applied Problems subscale for a pre-K specific subset of items (Weiland et al., 2011).

Procedures

As stated, this study evaluated a continuation of a TRIAD implementation (with pre-K procedures and analyses reported in Clements et al., 2011; Sarama et al., 2010). The same staff conducted teacher training, classroom observations, and child assessments and thus had substantial experience; however, all assessors were recertified on their respective instruments (e.g., they submit a video of a refresher session for evaluation and feedback until they have an errorless administration).

Teachers taught their assigned curriculum, the kindergarten level of the Investigations program (*Investigations*, 2008). They attended the professional development sessions in August, September, October, November, December, January, and March. They were encouraged during the TRIAD professional development sessions to use formative assessment based on the learning trajectories to modify activities, including omitting unnecessary instruction. In these sessions, teachers expressed the view that district rules required that they follow curricular schedules. They argued that these factors militated against their implementing the TRIAD instructional practices, especially curricular contraction.

During the pre-kindergarten year, we conducted COEMET classroom observation in every classroom. During the kindergarten year, the number of teachers expanded from 106 to 275 teachers. Because constraints on our resources prohibited observations in this many classrooms, we blocked the classrooms according to the number of study children in the classroom and sampled disproportionately. This allowed us to get a representative sample by covering a high proportion of students while visiting a much lower proportion of classrooms. Specifically, we compiled information on how many children from the original sample were in each classroom. We blocked by classrooms with a relatively large number of students (9 to 15), a medium number of students (5–8), a small number of students (2–4), and by classrooms with only one TRIAD student. We divided each of these four blocks into the three conditions (TRIAD-FT, TRIAD-NFT, and control), and conducted COEMET observations, across conditions, in all the classrooms with a large number of students, in most of the classrooms with a medium number, in a smaller proportion of those with two to four students per classroom, and in an even smaller proportion of those classrooms with only one participating student (125 classrooms in all). Finally, child assessors administered the REMA in the spring.

Data Analysis

Research questions related to both treatment group differences and mediation were answered with HLM using HLM6 (Raudenbush, Bryk, Cheong, & Congdon, 2004). Analyses included the use of pre-K scores (see Clements et al., 2011) as covariates at both the child and the mean aggregated at the school level. The full model for the estimation of differences between the TRIAD-FT and TRIAD-NFT condition from control, as well as the value added model comparing TRIAD-FT and TRIAD-NFT controlling for both pretest and posttest is displayed in the appendix.

RESULTS

Research questions are addressed in several sections. First, we examine the effects of the TRIAD interventions, compared to the control condition, across pre-K and kindergarten. Both ITT and TOT analyses are presented. Second, we measure just the effects of the TRIAD-FT during the kindergarten year by comparing the TRIAD-FT and TRIAD-NFT interventions. In both these sections, we assess potential moderators. Finally, we test mediational hypotheses involving the classroom observations.

Effects of TRIAD Over 2 Years

Main Effects. Table 2 presents the means and standard deviations of mathematics (REMA) scores by treatment condition and time point. ITT analyses were conducted on all available children, regardless of whether they had switched classrooms, schools, or condition. These analyses assessed the effect of the TRIAD interventions across two school years, using the pre-K pretest mathematics achievement score as a covariate at both the child and school levels and the end-of-kindergarten mathematics achievement scores as the dependent variable. The pretest covariate at the child level was entered group-mean centered. Further, both site and a blocking identifier are included at the school level to control for the impact of blocking and the slight variation in fourth grade math scores on which they are based (Kirk, 1982). All other predictors in the model are entered grand-mean centered. Because there was attrition, we performed post hoc power analysis; 29 children per school, 42 schools ($\rho = .25$, $r^2 = .60$), and power of .80 provided a minimal detectable effect size of .32 (Liu, Spybrook, Congdon, & Raudenbush, 2011). A two-level HLM analyses revealed a significant difference in kindergarten child outcomes in mathematics achievement between the TRIAD-FT group and the control group ($\beta = .225$; $SE = .059$, $p = .001$, $g = .34$). A significant difference was also found between the TRIAD-NFT and control conditions ($\beta = .134$; $SE = .058$, $p = .028$, $g = .21$). The difference between the TRIAD-FT and TRIAD-NFT conditions did not reach traditional levels of statistical significance ($\beta = .091$, $SE = .059$, $p = .132$, $g = .14$).

TOT analyses. All remaining analyses involved TOT; that is, they were conducted on only those children who stayed within their randomized condition and school and who completed all components of the assessment at every time point. A post hoc power analysis with a remaining average of 22 children per school, 42 schools ($\rho = .25$, $r^2 = .60$), and power of .80 provided a minimal detectable effect size of .33. Pretest covariates were entered at both the child and school levels.

Moderation. We also tested whether gender, class size, disability status, ethnicity, socioeconomic status, and site moderated experimental effects. Prerandomization moderators were entered into the model as dichotomized indicators wherein male was the comparative group for gender and the African American ethnic group was compared against all other ethnic groups (e.g., African American = 1, all other ethnic groups = 0, following our earlier results; Clements et al., 2011). There was no significant main effect for gender or disability status. There was a significant main effect for African American compared to other groups. Across research groups, children identifying as African American scored lower on the assessment of mathematics achievement at the end of kindergarten controlling for pretest scores as compared to children identifying with other racial/ethnic groups. The results for these comparisons are displayed in Table 3.

There were no significant interactions with treatment for gender or disability status. Significant interactions with the TRIAD-FT for the TRIAD-NFT and control group comparisons were found for children identifying as African American compared to children identifying with other ethnic backgrounds. The interaction with the TRIAD-NFT condition compared to control was not significant. Children identifying as African American within the TRIAD-FT group demonstrated significantly higher mathematics outcome scores than children identifying as African American within the TRIAD-NFT and control conditions (see Figure 3).

At the school level, the TRIAD-FT and TRIAD-NFT groups were each significantly different from the control group (TRIAD-FT, $g = .38$; TRIAD-NFT, $g = .32$). TRIAD-FT

Table 2. Means and standard deviations of mathematics outcome scores by treatment condition and time point

Condition by Time Point	Overall	Female	Male	AA	Other
TRIAD-FT	$N = 345$	$n = 185$	$n = 160$	$n = 202$	$n = 143$
Pre-K pretest	37.56	38.07	36.97	36.92	38.47
	(5.66)	(5.65)	(5.62)	(4.74)	(6.66)
Pre-K posttest	47.44	47.87	46.95	46.95	48.14
	(4.31)	(4.06)	(4.54)	(4.00)	(4.64)
Kindergarten	53.22	53.54	52.86	52.13	54.76
	(4.68)	(4.80)	(4.52)	(4.18)	(4.92)
TRIAD-NFT	$N = 333$	$n = 174$	$n = 159$	$n = 190$	$n = 143$
Pre-K pretest	37.15	37.67	36.57	36.38	38.17
	(6.06)	(6.12)	(5.97)	(6.04)	(5.96)
Pre-K posttest	47.01	47.35	46.64	46.16	48.14
	(4.51)	(4.10)	(4.91)	(4.43)	(4.39)
Kindergarten	52.64	53.13	52.09	51.41	54.26
	(4.44)	(4.32)	(4.52)	(4.20)	(4.25)
Control	$N = 285$	$n = 154$	$n = 131$	$n = 147$	$n = 138$
Pre-K pretest	37.81	37.64	38.01	37.84	37.78
	(5.59)	(6.06)	(5.01)	(4.99)	(6.19)
Pre-K posttest	44.15	44.16	44.13	42.96	45.42
	(5.09)	(5.12)	(5.09)	(5.08)	(4.82)
Kindergarten	51.61	51.22	52.06	50.18	53.12
	(4.69)	(4.43)	(4.96)	(4.49)	(4.44)

Note. Rasch scores converted to a T score ($M = 50$, $SD = 10$). Average scores by prerandomization moderators of gender and ethnicity are also displayed. TRIAD-NFT = TRIAD intervention in pre-K only; TRIAD-FT = TRIAD intervention in pre-K and follow through in kindergarten; AA = African American.

Table 3. Hierarchical linear models final fixed effects model outcomes and variance components for mathematics kindergarten posttest scores, with pre-K scores as covariates at both levels comparing each treatment group

TRIAD-NFT and TRIAD-FT Compared to Control				TRIAD-FT and Control Compared to TRIAD-NFT			
	Coeff	*SE*	*p*		Coeff	*SE*	*p*
Intercept	–1.114**	.020	.000	Intercept	–1.114**	.020	.000
Level 1 (Child)				Level 1 (Child)			
Pretest	.369**	.021	.000	Pretest	.369**	.021	.000
Gender (Male)	.018	.033	.584	Gender (Male)	.018	.033	.584
IEP Status	–.106	.059	.074	IEP Status	–.106	.059	.074
AA	–.233**	.037	.000	AA	–.233**	.037	.000
Level 2 (School)				Level 2 (School)			
Pretest Aggregate	.565**	.086	.000	Pretest Aggregate	.565**	.086	.000
Block	.009	.009	.346	Block	.009	.009	.346
TRIAD-FT	.251**	.048	.000	TRIAD-FT	.038	.048	.436
TRIAD-NFT	.213**	.048	.000	Control	–.213**	.048	.000
F/RL%	.002	.004	.651	F/RL%	.000	.004	.910
F/RL% × TRIAD-NFT	–.001	.005	.805	F/RL% × Control	.001	.005	.805
F/RL% × TRIAD-FT	–.007	.004	.092	F/RL% × TRIAD-FT	–.006	.004	.160
Site (Buffalo)	–.399**	.084	.000	Site (Buffalo)	–.308**	.080	.001
Site × TRIAD-FT	–.022	.123	.860	Site × TRIAD-FT	–.113	.120	.355
Site × TRIAD-NFT	.091	.117	.443	Site × Control	–.091	.117	.443

Class Size	–.005	.011	.651	Class Size	.005	.014	.735
Class Size × TRIAD-FT	.009	.018	.587	Class Size × TRIAD-FT	–.008	.021	.727
Class Size × TRIAD-NFT	.002	.021	.916	Class Size × Control	–.009	.018	.587
Interactions							
Gender × TRIAD-FT	–.095	.081	.242	Gender × TRIAD-FT	.027	.078	.727
Gender × TRIAD-NFT	–.122	.082	.135	Gender × Control	.122	.082	.135
IEP × TRIAD-FT	–.072	.146	.623	IEP × TRIAD-FT	–.077	.145	.594
IEP × TRIAD-NFT	–.005	.132	.970	IEP × Control	–.005	.142	.970
AA × TRIAD-FT	.366**	.096	.000	AA × TRIAD-FT	.243**	.093	.009
AA × TRIAD-NFT	.123	.088	.166	AA × Control	–.123	.089	.166
	Random Effect	*SD*	*p*				
Level 1	.249	.499					
Level 2	.003	.056	.001				

Note. TRIAD-NFT = TRIAD intervention in pre-K only; TRIAD-FT = TRIAD intervention in pre-K and follow through in kindergarten; IEP = Individualized Education Plan; AA = African American; F/RL% = school-level percentage of children receiving either free or reduced lunch.

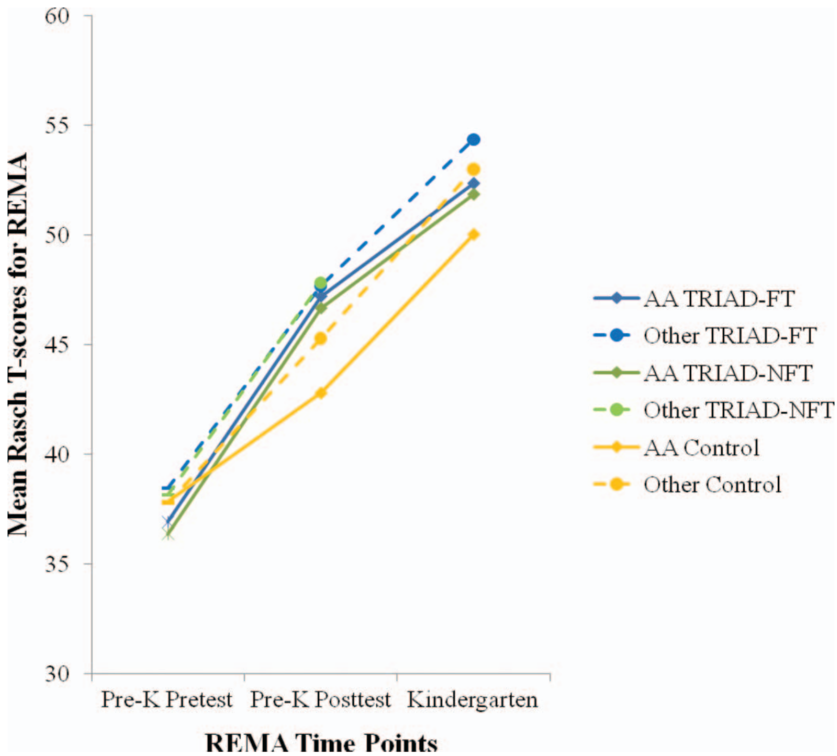


Figure 3. Comparison of African American (AA) and other racial/ethnic groups. *Note.* Analyses controlled for pretest only. TRIAD-FT = TRIAD condition with follow-through intervention; TRIAD-NFT = TRIAD condition pre-K intervention only; REMA = Research-based Elementary Math Assessment.

and TRIAD-NFT were not significantly different from one another ($g = .06$). Site had a significant impact on outcomes indicating that children within the Buffalo site scored lower overall. Treatment groups, however, did not differ significantly by site. The school aggregated pre-K pretest was also a significant predictor of kindergarten outcomes. Of the control variables included at the school level, percentage of free/reduced lunch, the blocking identifier, and school aggregated class-size main effect and interactions, none demonstrated a significant impact on child achievement.

Effects of the TRIAD-FT in the Kindergarten Year

Main Effect. As the TRIAD-FT and TRIAD-NFT groups received the same intervention during the 1st year of the intervention, a two-level HLM model was computed to determine the "value added" of the additional training received by the kindergarten teachers within the TRIAD-FT condition relative to pre-K only experience of the TRIAD-NFT group. The major distinction between these models is the addition of the pre-K posttest as a covariate at both the child and school levels. All previous moderators and control variables were also carried through this analysis. The TRIAD-FT group scored higher than the TRIAD-NFT, but again this difference was not significant (TRIAD-FT: $\beta = .049$, $SE = .048$; $p = .312$, $g = .07$).

Moderators. Regarding moderators, gender (male) demonstrated neither a significant main effect ($\beta = .033$, $SE = .026$, $p = .205$) nor interaction ($\beta = .062$, $SE = .062$, $p = .314$). Disability status also evinced neither a significant main effect ($\beta = -.031$, $SE = .047$, $p = .509$) nor interaction ($\beta = -.028$, $SE = .114$, $p = .808$). A significant main effect was found for African American, however ($\beta = -.149$, $SE = .032$, $p = .000$), whereby children who identify as African American demonstrated lower mathematics achievement scores at the end of kindergarten controlling for both pretest and posttest as compared to children identifying as a member of another ethnic/racial grouping. Further, a significant interaction with the TRIAD-FT condition was found. African American children in the TRIAD-FT condition scored higher than African American children in the TRIAD-NFT condition ($\beta = .169$, $SE = .078$, $p = .031$). At the school level, only the main effect for site was significant such that children within the Buffalo site scored lower overall as compared to the Boston site ($\beta = -.237$, $SE = .081$, $p = .007$).

Mediation

As the intervention conditions include a predominant focus on training and instruction at the teacher level, we were also interested in examining questions regarding the mediational role of instructional processes on child-level mathematics achievement at the end of kindergarten. These analyses were conducted with school-aggregated COEMET variables as the average number of teachers per school was 2.7 and the intervention was applied at the school level. Specifically, the mediational hypotheses assess whether the indirect effect of the treatment group through aspects of the COEMET on child kindergarten math achievement differs significantly from zero. Utilizing the notation (*m*) and (*y*) to reference the mediator and the outcome (Pituch, Stapleton, & Kang, 2006), respectively, the standard regression equation for the impact of the treatment group on the mediator can be

$$M_j = \gamma_{0\text{o}(m)} + aX_j + u_{\text{o}j(m)},$$

where M_j and X_j represent the school-level mediator and treatment condition and a represents the impact of the treatment condition on the mediator. The intercept and residual for the equation are estimated as $\gamma_{0\text{o}(m)}$ and $u_{\text{o}j(m)}$, respectively. To estimate the impact of the mediator on child achievement posttest at the end of kindergarten, controlling for the pretest, a two-level HLM was constructed as

$$Y_{ij} = \beta_{0j(y)} + \beta_{1j(y)} + r_{ij(y)}.$$

where Y_{ij} is the child outcome at kindergarten posttest, $\beta_{0j(y)}$ is the intercept, $\beta_{1j(y)}$ is the slope of the REMA pretest, and $r_{ij(y)}$is the residual for the equation. The school-level equation for the impact of each mediator on the outcome is

$$\beta_{0j(y)} = y_{0\text{o}(y)} + c'X_j + bM_j + u_{0j(y)}.$$

where the effect of the mediator on the outcome, *b,* is estimated controlling for the effect of treatment, and c' is the direct effect of treatment on the outcome controlling for the mediator.

The indirect effect is represented by the cross-product (*ab*) for the *a* and *b* unstandardized regression coeffecients (Preacher & Hayes, 2008). We calculated 95% confidence intervals for the *ab* product by submitting the unstandardized regression coefficients and their standard errors to the PRODCLIN program (MacKinnon, Fritz, Williams, & Lockwood, 2007). Based on the distribution of the *ab* product, the PRODCLIN program computes the asymmetric confidence intervals. The use of this procedure has been found to demonstrate both high power and low Type 1 error rates (MacKinnon, Lockwood, & Williams, 2004). Mediators were tested both individually and within a multiple mediator model. The primary purpose for the multiple mediator analysis is to account for the relationship between the mediators in the prediction of the outcome (Preacher & Hayes, 2008). The COEMET variables measured at the end of kindergarten, however, evinced both small and insignificant relationships amongst the measured components, none reaching above $r^2 =$ .30. We tested the individual mediator results against the multiple mediator model and found the pattern of significance to be the same.

For COEMET variables measured at kindergarten only, a significant indirect effect was found through the number of SMAs for the TRIAD-FT group as compared to the control group on child outcome scores at the end of kindergarten $\{a = 1.92\ (.859);\ b = .042\ (.017)\}$. This indirect effect accounted for 25% of the total effect of the TRIAD-FT group as compared to control on child outcomes. In addition, a significant indirect effect for the classroom culture subscale was found for the comparison of the TRIAD-NFT and TRIAD-FT conditions $\{a = 3.26\ (1.15);\ b = .043\ (.015)\}$. This indirect effect accounted for 73% of the total effect. Neither the SMA subscale nor the time on task acted in a significant way to transmit the effect of any condition on child outcomes at the end of kindergarten. Controlling for pre-K posttest as well, the value added comparison of TRIAD-FT and TRIAD-NFT, the Classroom Culture subscale continued to mediate the relationship with child outcome scores $\{a = 3.26\ (1.15);\ b = .015\ (.010)\}$, accounting for 89% of the total effect. These results are displayed in Table 4.

For COEMET variables representing the sum of the pre-K and kindergarten measurement time points, the number of SMAs continued to demonstrate a significant indirect effect on the impact of TRIAD-NFT $\{a = 4.56\ (.1.73);\ b = .032\ (.007)\}$ and TRIAD-FT $\{a = 6.98\ (1.87);\ b = .286\ (.008)\}$, each relative to control, on child outcomes at kindergarten. The indirect effect accounted for 67% of the total effect for the comparison of TRIAD-NFT to control and 63% for the comparison of TRIAD-FT to control. The comparison between TRIAD-FT and TRIAD-NFT did not demonstrate a significant indirect effect through the number of specific math activities on child outcomes. The Classroom Culture subscale demonstrated an indirect effect on the impact of TRIAD-FT relative to TRIAD-NFT on child outcomes $\{a = 6.04\ (2.97);\ b = .024\ (.008)\}$, accounting for 89% of the total effect. The TRIAD-FT group led to greater focus on mathematics, which led to an increase in child outcome scores for the TRIAD-FT group relative to TRIAD-NFT. When controlling for both pretest and posttest in the value-added comparison of the TRIAD-NFT and TRIAD-FT group, the sum of the pre-K and kindergarten COEMET scores did not demonstrate a significant indirect effect on child outcomes. These results are displayed in Table 5.

DISCUSSION

Scale-up research must examine not only immediate effects but also the persistence of effects. This study used a cluster randomized trial design to evaluate the effectiveness of an implementation of a research-based model for scaling up educational interventions,

Table 4. Indirect effects with 95% confidence intervals covarying pretest scores

Kindergarten COEMET Component and Condition Comparison	*ab* Point Estimate	Pretest Covariate		*ab* Point Estimate	Pretest and Posttest Covariates – Value Added	
		Lower	Upper		Lower	Upper
Classroom Culture subscore						
TRIAD-FT vs. control	.0887	–.0053	.2252			
TRIAD-NFT vs. control	.0027	–.1059	.1126			
TRIAD-FT vs. TRIAD-NFT	.1392*	.0238	.3024	.0476*	.0153	.1339
SMA subscore						
TRIAD-FT vs. control	.0048	–.0593	.0378			
TRIAD-NFT vs. control	.0021	–.0367	.0471			
TRIAD-FT vs. TRIAD-NFT	.0095	–.0411	.0797	.0041	–.0027	.0442
Total no. of SMAs (M)						
TRIAD-FT vs. control	.0813*	.0022	.2005			
TRIAD-NFT vs. control	.0418	–.0565	.1596			
TRIAD-FT vs. TRIAD-NFT	.0621	–.0410	.1955	.0337	–.0094	.1036
Time on Task (min/day)						
TRIAD-FT vs. control	.0011	–.0240	.0201			
TRIAD-NFT vs. control	.0009	–.0189	.0223			
TRIAD-FT vs. TRIAD-NFT	.0011	–.0026	.0025	.0004	–.0026	.0025

Note. AA = African American; TRIAD-FT = TRIAD condition with follow-through intervention; TRIAD-NFT = TRIAD condition pre-K intervention only; SMA = specific mathematics activity.
*$p < .05$.

Table 5. Indirect effects with 95% confidence intervals covarying pretest and posttest scores

Pre-K + Kindergarten COEMET Component and Condition Comparison	ab Point Estimate	95% Confidence Intervals: Pretest Covariate Lower	Pretest Covariate Upper	ab Point Estimate	Pretest and Posttest Covariates – Value Added Lower	Pretest and Posttest Covariates – Value Added Upper
Classroom culture subscore						
TRIAD-FT vs. control	.1215	–.0535	.3270			
TRIAD-NFT vs. control	.0617	–.0601	.2221			
TRIAD-FT vs. TRIAD-NFT	.1452*	.0028	.3477	.0316	–.1326	.0444
SMA subscore						
TRIAD-FT vs. control	.0581	–.1689	.0150			
TRIAD-NFT vs. control	.0140	–.0859	.1285			
TRIAD-FT vs. TRIAD-NFT	.0456	–.1844	.0397	.0226	–.1427	.0520
Total no. of SMAs (M)						
TRIAD-FT vs. control	1.999*	.9486	3.058			
TRIAD-NFT vs. control	.1465*	.0336	.2899			
TRIAD-FT vs. TRIAD-NFT	.0826	–.0467	.2297	.0215	–.0197	.08942
Time on task (min/day)						
TRIAD-FT vs. control	.0002	–.0148	.0154			
TRIAD-NFT vs. control	.0061	–.0241	.0456			
TRIAD-FT vs. TRIAD-NFT	.0005	–.0206	.0189	.0009	–.0324	.0356

Note. AA = African American; TRIAD-FT = TRIAD condition with follow-through intervention; TRIAD-NFT = TRIAD condition pre-K intervention only; SMA = specific mathematics activity.

*$p < .05$.

focusing on the persistence of effects of two interventions based on the TRIAD model of scale up (Sarama et al., 2008). The interventions were identical at pre-K, but only one included the follow-through component, that is, an intervention based on the TRIAD model in the kindergarten year. We discuss the results in four categories. First, we discuss the effects of the two interventions on mathematics achievement at the end of kindergarten, with ITT analyses that include all children in their original, randomly assigned groups, regardless of how much of the intervention they experienced. Second, we discuss the same effects with TOT analyses, which provide information on those children who remained in their treatment groups. Third, we report whether these effects differed for different groups by examining several possible moderators of the effects of the interventions. Fourth, we examine which, if any, measured components of the quantity and quality of the classroom environments and teachers mediated the effects of the TRIAD interventions.

First, addressing effects on children's mathematics achievement at the end of their kindergarten year, the ITT analyses assessed the effect of the TRIAD interventions over the 2 school years. Results showed a significant difference between the TRIAD-FT group and the control group ($g = .34$). The TRIAD-NFT group was also significantly different the control group ($g = .21$). The two treatment conditions were not statistically significantly different ($g = .14$). Even under the effects of dilution through mobility and attrition, the TRIAD-FT interventions and, to a lesser extent, the TRIAD-NFT intervention, demonstrated significantly higher scores than the business-as-usual control condition.

Second, TOT analyses were conducted on all children who remained in their randomized condition and school throughout the 2 years and completed all assessments. Consistent with the results from pre-K (in which the experimental interventions were identical; Clements et al., 2011), both TRIAD groups, TRIAD-FT and TRIAD-NFT, were statistically significantly higher in mathematics achievement than the control group. Consistent with the hypotheses of this study, the effect size for the TRIAD-FT group ($g = .38$) was greater than the effect size for the TRIAD-NFT group ($g = .32$). However, the two TRIAD groups were not significantly different at the end of the kindergarten year ($g = .06$). Similarly, using both pre-K pretest and pre-K posttest as covariates, the TRIAD-FT group scored higher than the TRIAD-NFT, but again the difference did not reach traditional levels of statistical significance ($g = .07$). The anticipated sample size was larger than the realized sample size, and thus the power was lower than we initially designed. Given the lower power of this longitudinal study and the tendency for the TRIAD-FT group to outperform the TRIAD-NFT group, future research should continue to investigate these comparisons.

The TRIAD interventions were, therefore, successful over the pre-K and kindergarten years. Previous studies reported effect sizes for the pre-K TRIAD intervention ranging from .62 (Sarama et al., 2008) to .72 (Clements et al., 2011, recalled this was a report of this same study at pre-K). Extending the TRIAD intervention through kindergarten lessened these effects, more so for the non-follow-through group than for the follow-through group. Therefore, the evidence supports the persistence of the effects of this implementation of the TRIAD model, at least considering the TOT analyses. The evidence also suggests that the implementation of follow-through (arguably one more extensive and intense than the present implementation) may be important to maintaining the trajectory of learning enhanced by the pre-K intervention.

These findings can be interpreted in the context of our theoretical framework. Several factors that we posited as deleterious to the persistence of preschool interventions were demonstratively present at kindergarten in both school districts involved in this study (the presence of other factors was likely). One adopted a new edition of the mathematics curriculum the same year that TRIAD intervention was implemented in kindergarten.

Teachers in both districts were required to follow their curricula strictly. Further, the entire first semester of that curriculum covers only concepts and skills already developed in the *Building Blocks* pre-K program. This may have attenuated the effect of the follow-through intervention and thus led to the lack of statistically significant differences between the TRIAD only group and the TRIAD-FT group.

Third, we examined potential moderators of gender and ethnic group and, at the school level, percentage of students who receive free or reduced lunch, and site. Only one interacted with treatment group statistically and practically significantly. Multiple analyses revealed significant interactions between the TRIAD-FT intervention and children identified as African American versus children who identified with any other ethnic group(s). African American children within the TRIAD-FT group scored significantly better on child kindergarten outcomes than African American children in the TRIAD-NFT and control groups. The additional training provided to kindergarten teachers in the TRIAD-FT intervention, expanding on the high-quality preschool math, engendered even greater gains in mathematics achievement for African American children than the preschool experience alone. This supports the value of the TRIAD-FT intervention in the kindergarten year, possibly especially for groups underrepresented in mathematics achievement. It may be that focusing on learning trajectories centers teachers' attention on mathematical concepts and processes and, more so, on children's thinking and learning, rather than on each child as a member of an ethnic group (wherein biases may negatively affect school experiences, as reported in U.S. Department of Health and Human Services—Administration for Children and Families, 2010).

Fourth, we examined what components of the classroom observation instrument mediated the effects. A statistically significant portion of the effect of the TRIAD-FT intervention on child outcomes, relative to the control condition, is transmitted through the Number of Specific Math Activities observed. This is consistent with analyses on the pre-K intervention implemented with these same children (Clements et al., 2011). That is, the number of distinct mathematics activities in which children engage is more important than total time on task in supporting their learning of mathematics (cf. Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2005). There are several possible explanations for this finding. For example, it may be that developmental constraints, such as limits on attention, result in diminishing returns in longer activities. A more cogent explanation may be that children learn more from a variety of activities emphasizing the same level of thinking, as they may learn concepts more readily from generalizing mathematics structures from different problematic situations that require the same mathematical concepts and processes (e.g., mental actions-on-objects) for their solution. Further, such multiple situations may create a greater number of cognitive paths for retrieval of these concepts and processes. A caveat is that both the findings and hypothesizing are post hoc, and should be tested in future research. Also consistent with the pre-K results, the total time on task did not statistically significantly mediate the effects of the TRIAD follow-through intervention; rather, it was the number of activities that did so.

The classroom culture component of mathematics teaching also mediated the effect of the TRIAD follow-through intervention compared to the TRIAD-NFT condition, again consistent with the pre-K intervention implemented with these same children (Clements et al., 2011). Thus, exposure to the follow-through intervention led to a greater focus on mathematics in the classroom, which in turn led to an increase in child outcome scores. The impact of classroom culture on math scores is consistent with the literature supporting the connection between academic performance and general features of the classroom, including signs of mathematical activity, teachers who are knowledgeable and enthusiastic about mathematics, and teachers who interact with and respond to children frequently (Clarke

& Clarke, 2004; Clements & Sarama, 2007b; Fraivillig, Murphy, & Fuson, 1999; Sawada et al., 2002; S. P. Wright, Horn, & Sanders, 1997). These features, as well as the greater numbers of the specific mathematics activities, may be a result of altered teachers' beliefs and dispositions, with TRIAD-FT teachers becoming more positive and more engaged. A caveat is that the portion of the effect accounted for by these and other components of the classroom observation instrument were small.

Implications

Educational theory, research, and practice need transferable, practical examples of scale up (McDonald et al., 2006) that are empirically supported to increase children's knowledge (Borman, 2007). Results here and in previous publications support the effectiveness of the TRIAD model. The small to moderate effect sizes for the persistence of effects, .32 for the TRIAD group (.21 for the ITT analysis) without follow-through compared to the control group and .38 for the TRIAD group with follow-through (.34 for the ITT analysis) remain notable in comparison to effect sizes of comprehensive reform programs, including multitiered teacher support, sustained professional development, and in-class coaching, which obtain effect sizes such as .24 only with great effort (Balfanz, Mac Iver, & Byrnes, 2006). Thus, practitioners might apply the specific instantiation of the model as described here, including a pre-K curriculum built upon research-based learning trajectories and follow-up interventions with kindergarten teachers using such learning trajectories for formative assessment, allowing more extensive modifications such as curricular contraction. Such interventions are potentially critical. The best predictor of a successful academic career is early mastery of literacy and mathematical concepts and skills (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Duncan, Claessens, & Engel, 2004; Paris et al., 2006; Stevenson & Newman, 1986). This is especially important, because research suggests that teacher quality affects children's learning more in low-resource than in high-resource schools, with larger effects on mathematics than reading (Nye, Konstantopoulos, & Hedges, 2004). Further, students from low-resource communities benefit more relative to children from higher resource communities from the same "dose" of school instruction (Raudenbush, 2009). Thus, interventions based on models such as the TRIAD follow-through model, which include substantial professional development and coaching, may be especially effective in low-resource schools such as those in this study.

The success of this and previous implementations of the TRIAD model have three additional implications for research. First, although the research design cannot identify which components of the TRIAD model and its instantiation in the present study, such disambiguation would be a worthwhile goal of future studies. Second, researchers might attempt to apply the model to other grade levels and other subject matter emphases, as TRIAD was designed as a general model of scaling up successful interventions. (The effects of the follow-through treatment suggest at least that the TRIAD model is not yoked to a single curriculum.) TRIAD's 10 research-based guidelines are consistent with, but more elaborated than, subsequent generalizations from the empirical corpus (Pellegrino, 2007).

The third implication for research is that the present study is another in a growing body of empirical research that supports learning trajectories, including evaluations of curricula based on the same trajectories as this study (Clements & Sarama, 2007c, 2008a; Sarama et al., 2008), elementary curricula based on related trajectories (e.g., Math Expressions in Agodini & Harris, 2010), as well as studies of teaching (Wood & Frid, 2005) and professional development projects (Bright, Bowman, & Vacc, 1997; Clarke et al., 2002; R. J. Wright, Martland, Stafford, & Stanger, 2002). The effect of the TRIAD follow-through intervention is significant in this regard. Whereas the intervention in pre-K included a new curriculum with multiple components, the kindergarten intervention included only one supplemental curriculum component, the *Building Blocks* software. The core of the follow-through intervention was teachers' use of the learning trajectories, in a minor way through the software, but more substantially through modification of the instructional activities from their regular curriculum. The teachers' use of the learning trajectories was not intrinsically connected to their existing mathematics curriculum but was a general approach to formative assessment that has been shown to be effective in multiple curricula and situations (e.g., Foorman et al., 2007; National Mathematics Advisory Panel, 2008). A caveat is that observations of how the teachers implemented their knowledge of learning trajectories was limited and related to different explanations, such as increased enthusiasm and attention to mathematics, which might also have contributed to the effects.

Finally, our findings have two additional implications for practice. First, multiple studies have reported that preschool gains "fade" (Fish, 2003; Natriello et al., 1990; Preschool Curriculum Evaluation Research Consortium, 2008; Turner & Ritter, 2004; U.S. Department of Health and Human Services—Administration for Children and Families, 2010). The results reported here indicate that the effects of the TRIAD implementation of the *Building Blocks* pre-K mathematics program can persist, with significantly higher average scores for both TRIAD groups, with and without follow through, at the end of the kindergarten year. Only the intervention with follow-through produced effects that persisted as measurable and statistically significant for students regardless of their movement between classrooms, schools, and thus conditions. This evidence supports claims that educators must continue to engage children in interesting and challenging mathematics to ensure academic success (Brooks-Gunn, 2003; Paris et al., 2006).

The ostensible "fading" of preschool effects is often reported without adequate attention to the follow-up—more frequently, the *lack* of follow-up—planned and implemented for these children. We designed and evaluated the effectiveness of TRIAD's follow-through intervention, testing our hypothesis that such follow-through is the "missing piece" in many early interventions whose longitudinal evaluations have found less positive effects. We believe this mistakenly treats initial effects of interventions as independent of the future school contexts. That is, they theoretically reify the treatment effect as an entity that should persist unless it is "weak" or evanescent, susceptible to fading. Instead, we believe children's trajectories must be studied as the children experience different educational courses. Treatment effects are relative, both in contrasting experimental and control groups and, longitudinally, to the nature of educational experiences these groups subsequently receive.

Thus, especially given the cumulative effect of high-quality teaching (Sanders & Rivers, 1996), it is essential that researchers continue to investigate the effects of early and follow-through interventions. This statement is, perhaps, easier to agree with than to implement. In the TRIAD-FT intervention presented here, multiple factors impeded implementation, including teacher's views that district rules required following schedules and would not allow such substantive formative assessment strategies as curriculum contraction. Researchers and practitioners would benefit from addressing such problems collaboratively.

Second, this study substantiated the findings from the original evaluation of these children in their pre-K year (Clements et al., 2009) indicating that the TRIAD implementation was particularly successful for children who identified themselves as African American. Although African American children continued to lag behind non–African American children in all conditions, the TRIAD-FT intervention helped them narrow that achievement gap through their kindergarten year. A high-quality, consistent mathematics education can make a demonstrative and consistent positive impact on the educational attainment of African American children in the pre-K and kindergarten years compared to traditional instruction. This is particularly important for African American children, as they are likely to suffer from low expectations for achievement above and beyond the impacts of poverty (National Mathematics Advisory Panel, 2008).

Third and finally, the TRIAD-FT intervention's effect was partially due to the increase in the positive classroom cultures teachers develop and through the number of specific math activities that they implement. Thus, interventions such as these can help engender a greater focus on mathematics, which in turn can help increase students' mathematics achievement. As other work has shown (Carpenter, Fennema, Peterson, & Carey, 1988; Clements et al., 2011; Jacobs, Franke, Carpenter, Levi, & Battey, 2001; National Research Council, 2009), helping primary teachers' gain additional knowledge of mathematics, students' thinking and learning about mathematics, and how instructional tasks can be designed and modified—that is, the three components of learning trajectories—has a measurable, positive effect on their students' achievement.

REFERENCES

Agodini, R., & Harris, B. (2010). An experimental evaluation of four early elementary school math curricula. *Journal of Research on Educational Effectiveness*, *3*, 199–253.

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental dynamics of math performance from pre-school to Grade 2. *Journal of Educational Psychology*, *96*, 699–713.

Baker, E. L. (2007). Principles for scaling up: Choosing, measuring effects, and promoting the widespread use of educational innovation. In B. Schneider & S.-K. McDonald (Eds.), *Scale up in education, Volume I: Ideas in principle* (pp. 37–54). Lanham, MD: Rowman & Littlefield.

Balfanz, R., Mac Iver, D. J., & Byrnes, V. (2006). The implementation and impact of evidence-based mathematics reforms in high-poverty middle schools: A multi-site, multi-year study. *Journal for Research in Mathematics Education*, *37*, 33–64.

Bennett, N., Desforges, C., Cockburn, A., & Wilkinson, B. (1984). *The quality of pupil learning experiences*. Hillsdale, NJ: Erlbaum.

Berends, M., Kirby, S. N., Naftel, S., & McKelvey, C. (2001). *Implementation and performance in New American Schools: Three years into scale-up*. Santa Monica, CA: RAND Education.

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.

Bodilly, S. J. (1998). *Lessons from New American Schools' scale-up phase*. Santa Monica, CA: RAND Education.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.

Borman, G. D. (2007). Designing field trials of educational interventions. In B. Schneider & S.-K. McDonald (Eds.), *Scale up in education, Volume II: Issues in practice* (pp. 41–67). Lanham, MD: Rowman & Littlefield.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievment: A meta-analysis. *Review of Educational Research*, *73*, 125–230.

Bredekamp, S. (2004). Standards for preschool and kindergarten mathematics education. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics:*

Standards for early childhood mathematics education (pp. 77–82). Mahwah, NJ: Erlbaum.

Bright, G. W., Bowman, A. H., & Vacc, N. N. (1997). Teachers' frameworks for understanding children's mathematical thinking. In E. Pehkonen (Ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 105–112). Lahti, Finland: University of Helsinki.

Broberg, A. G., Wessels, H., Lamb, M. E., & Hwang, C. P. (1997). Effects of day care on the development of cognitive abilities in 8-year-olds: A longitudinal study. *Developmental Psychology*, *33*, 62–69.

Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report*, *17*(1), 3–14.

Carneiro, P., & Heckman, J. J. (2003). Human capital policy. In A. B. Krueger & J. J. Heckman (Eds.), *Inequality in American: What role for human capital policies?* (pp. 77–239). Cambridge, MA: MIT Press.

Carpenter, T. P., Fennema, E. H., Peterson, P. L., & Carey, D. A. (1988). Teacher's pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education*, *19*, 385–401.

Carpenter, T. P., & Franke, M. L. (2004). Cognitively guided instruction: Challenging the core of educational practice. In J. Glennan, K. Thomas, S. J. Bodilly, J. R. Galegher, & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 41–80). Santa Monica, CA: RAND Corporation.

Clarke, D. M., Cheeseman, J., Gervasoni, A., Gronn, D., Horne, M., McDonough, A., . . . Rowley, G. (2002). *Early Numeracy Research Project final report*. Department of Education, Employment and Training, the Catholic Education Office (Melbourne), and the Association of Independent Schools Victoria.

Clarke, D. M., & Clarke, B. A. (2004). Mathematics teaching in K–2: Painting a picture of challenging, supportive and effective classrooms. In R. Rubenstein & G. Bright (Eds.), *Perspectives on teaching mathematics: 66th Yearbook* (pp. 67–81). Reston, VA: National Council of Teachers of Mathematics.

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning*, *6*, 81–89.

Clements, D. H., & Sarama, J. (2007a). Building Blocks [Computer software]. Columbus, OH: SRA/McGraw-Hill.

Clements, D. H., & Sarama, J. (2007b). Early childhood mathematics learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 461–555). New York: Information Age Publishing.

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

Clements, D. H., & Sarama, J. (2008a). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*, 443–494.

Clements, D. H., & Sarama, J. (2008b, March). *Scaling-up interventions: The case of mathematics*. Paper presented at the American Educational Research Association, New York.

Clements, D. H., Sarama, J., & Liu, X. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-based Early Maths Assessment. *Educational Psychology*, *28*(4), 457–482.

Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York, NY: Routledge.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, *42*(2), 127–166.

Cohen, D. K. (1996). Rewarding teachers for student performance. In S. H. Fuhrman & J. A. O'Day (Eds.), *Rewards and reforms: Creating educational incentives that work* (pp. 61–112). San Francisco, CA: Jossey-Bass.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco, CA: Jossey-Bass.

Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade*. Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002125

Duncan, G. J., Claessens, A., & Engel, M. (2004). *The contributions of hard skills and socio-emotional behavior to school readiness*. Evanston, IL: Northwestern University

Elmore, R. F. (1996). Getting to scale with good educational practices. *Harvard Educational Review*, *66*, 1–25.

Fish, R. (2003). *Effects of attending prekindergarten on academic achievement* (Unpublished master's thesis). University of Buffalo, State University of New York, Buffalo, NY.

Fish, R. (2007). *Relationship between education intensity in kindergarten and grade 1 and the academic benefits of attending preschool* (Unpublished doctoral dissertation). niversity of Buffalo, State University of New York, Buffalo, NY.

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice*, *19*, 531–540.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication #231). Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, the National Implementation Research Network.

Foorman, B. R., Santi, K. L., & Berger, L. (2007). Scaling assessment-driven instruction using the Internet and handheld computers, Volume 2. In B. Schneider & S.-K. McDonald (Eds.), *Scale up in practice* (pp. 69–89). Lanham, MD: Rowan & Littlefield.

Fraivillig, J. L., Murphy, L. A., & Fuson, K. C. (1999). Advancing children's mathematical thinking in *Everyday Mathematics* classrooms. *Journal for Research in Mathematics Education*, *30*, 148–170.

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).

Glennan, J., Thomas K., Bodilly, S. J., Galegher, J. R., & Kerr, K. A. (Eds.). (2004). *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* Santa Monica, CA: RAND Corporation.

Gray, S. W., Ramsey, B. K., & Klaus, R. A. (1983). The early training project 1962–1980. In Consortium for Longitudinal Studies (Ed.), *As the twig is best ... Lasting effects of preschool programs* (pp. 33–69). Hillsdale, NJ: Erlbaum.

Griffin, S., Case, R., & Capodilupo, A. (1995). Teaching for understanding: The importance of the Central Conceptual Structures in the elementary mathematics curriculum. In A. McKeough, J. Lupart, & A. Marini (Eds.), *Teaching for transfer: Fostering generalization in learning* (pp. 121–151). Mahwah, NJ: Erlbaum.

Guskey, T. R. (Ed.). (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin.

Hall, G. E., & Hord, S. M. (2001). *Implementing change: Patterns, principles, and potholes*. Boston, MA: Allyn and Bacon.

Hiebert, J. C. (1999). Relationships between research and the NCTM Standards. *Journal for Research in Mathematics Education*, *30*, 3–19.

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. (1992). Critical introduction. In M. G. Fullan (Ed.), *Successful school improvement* (pp. 1–20). Philadelphia, PA: Open University Press.

Investigations in number, data, and space (2nd ed.). (2008). Upper Saddle River, NJ: Pearson Scott Foresman.

Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2001). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education*, *38*, 258–288.

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.

Kaser, J. S., Bourexis, P. S., Loucks-Horsley, S., & Raizen, S. A. (1999). *Enhancing program quality in science and mathematics*. Thousand Oaks, CA: Corwin.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Kirk, R. E. (1982). *Experimental design: Procedures for the social scientist*. Belmont, CA: Wadsworth.

Klingner, J. K., Ahwee, S., Pilonieta, P., & Menendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional Children*, *69*, 411–429.

Lawless, K. A., & Pellegrino, J. W. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*, 454–499.

Liu, X. F., Spybrook, J., Congdon, R., & Raudenbush, S. W. (2011). *Optimal design for multi-level and longitudinal research* (Ver. 2) [Computer software]. Lincolnwood, IL: Scientific Software.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, *39*, 384–389.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128.

Magnuson, K. A., & Waldfogel, J. (2005). Early childhood care and education: Effects on ethnic and racial gaps in school readiness. *The Future of Children*, *15*, 169–196.

McDonald, S.-K., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). Scaling-up exemplary interventions. *Educational Researcher*, *35*(3), 15–24.

Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly*, *21*, 313–331.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., & Smith, T. A. (2000). *TIMSS 1999 international mathematics report*. Boston, MA: The International Study Center, Boston College, Lynch School of Education.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

National Research Council. (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

National Research Council. (2009). *Mathematics in early childhood: Learning paths toward excellence and equity*. Washington, DC: National Academy Press.

Natriello, G., McDill, E. L., & Pallas, A. M. (1990). *Schooling disadvantaged children: Racing against catastrophe*. New York, NY: Teachers College Press.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*, 237–257.

Paris, S. G., Morrison, F. J., & Miller, K. F. (2006). Academic pathways from preschool through elementary school. In P. Alexander & P. Winne (Eds.), *Handbook of research in educational psychology* (pp. 61–85). Mahwah, NJ: Erlbaum.

Pellegrino, J. W. (2007). From early reading to high school mathematics: Matching case studies of four educational innovations against principles for effective scale up. In B. Schneider & S.-K. McDonald (Eds.), *Scale up in practice* (pp. 131–139). Lanham, MD: Rowan & Littlefield.

Pituch, K. A., Stapleton, L. M., & Kang, J. Y. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research*, *41*, 367–400.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness* (NCER 2008–2009). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education, U.S. Government Printing Office. Retrieved from http://ncer.ed.gov

Raudenbush, S. W. (2007). Designing field trials of educational interventions. In B. Schneider & S.-K. McDonald (Eds.), *Scale up in education, Volume II: Issues in practice* (pp. 23–40). Lanham, MD: Rowman & Littlefield.

Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal*, *45*, 206–230.

Raudenbush, S. W. (2009). The *Brown* legacy and the O'Connor challenge: Transforming schools in the images of children's potential. *Educational Researcher*, *38*, 169–180.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement* (Research Progress Report). Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.

Sarama, J., Clements, D. H., & Henry, J. J. (1998). Network of influences in an implementation of a mathematics curriculum innovation. *International Journal of Computers for Mathematical Learning*, *3*, 113–148.

Sarama, J., Clements, D. H., Starkey, P., Klein, A., & Wakeley, A. (2008). Scaling up the implementation of a pre-kindergarten mathematics curriculum: Teaching for understanding with trajectories and technologies. *Journal of Research on Educational Effectiveness*, *1*, 89–119.

Sarama, J., & DiBiase, A.-M. (2004). The professional development challenge in preschool mathematics. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 415–446). Mahwah, NJ: Lawrence Erlbaum Associates.

Sarama, J., Lange, A., Clements, D. H., & Wolfe, C. B. (in press). The impacts of an early mathematics curriculum on emerging literacy and language. *Early Childhood Research Quarterly*.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics*, *102*, 245–253.

Showers, B., Joyce, B., & Bennett, B. (1987). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. *Educational Leadership*, *45*(3), 77–87.

Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: Council of the Great City Schools.

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development*, *57*, 646–659.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2005). *The effective provision of pre-school education [EPPE] project: A longitudinal study funded by the DfEE (1997–2003)*. London, UK: EPPE Project, Institute of Education, University of London.

Turner, R. C., & Ritter, G. W. (2004, April). *Does the impact of preschool childcare on cognition and behavior persist throughout the elementary years?* Paper presented at the American Educational Research Association, San Diego, CA.

U.S. Department of Health and Human Services—Administration for Children and Families. (2010). *Head Start impact study. Final report*. Washington, DC: Office of Planning, Research and Evaluation, U.S. Department of Health and Human Services.

Weiland, C., Wolfe, C. B., Hurwitz, M., Clements, D. H., Sarama, J., & Yoshikawa, H. (2011, March-April). *Psychometrics and validation of the short form of an early mathematics assessment*. Paper presented at the Society for Research in Child Development, Montreal, Quebec, Canada.

Wolfe, B. (1991). *Effective practives in inservice education: An exploratory study of the perceptions of head start participants* (Unpublished doctoral dissertation). University of Wisconsin–Madison, Madison, WI.

Wood, K., & Frid, S. (2005). Early childhood numeracy in a multiage setting. *Mathematics Education Research Journal*, *16*(3), 80–99.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago: MESA Press.

Wright, R. J., Martland, J., Stafford, A. K., & Stanger, G. (2002). *Teaching number: Advancing children's skills and strategies*. London, UK: Paul Chapman/Sage.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*, 57–67.

APPENDIX

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\mathit{REMAPRE}) + \beta_{2j}(\mathit{Male}) + \beta_{3j}(\mathit{AA}) + \beta_{4j}(\mathit{IEP}) + e_{ij}$$

where

Y_{ij} is the mathematical competence of child i in school j;
β_{0j} is the mean outcome status adjusted for other predictors for children in school j;
β_{1j} is the slope of the REMA pretest (REMAPRE) of children in school j;
β_{2j} is the main effect for the dummy code indicator of Gender (Male) of children in school j;
β_{3j} is the main effect for the dummy code indicator of African American (AA) of children in school j;
β_{4j} is the main effect for the dummy code indicator of Individual Education Plan (IEP) of children in school j;
and e_{ij} is the residual (level – 1 random effect).

The equation for the school level is

$$\begin{aligned}\beta_{0j} = \gamma_{00} &+ \gamma_{01}(\mathit{School\ Aggregate\ pretest}) + \gamma_{02}\,(\mathit{Block}) + \gamma_{03}\,(\mathit{TRIADFT})\\ &+ \gamma_{04}\,(\mathit{TRIADNFT}) + \gamma_{05}\,(\mathit{FRL\%}\ ) + \gamma_{06}\,(\mathit{FRL\%} * \mathit{TRIADFT})\\ &+ \gamma_{07}\,(\mathit{FRL\%} * \mathit{TRIADNFT}) + \gamma_{08}\,(\mathit{Site}) + \gamma_{09}\,(\mathit{Site} * \mathit{TRIADFT})\\ &+ \gamma_{010}\,(\mathit{Site} * \mathit{TRIADNFT}) + \gamma_{011}\,(\mathit{Class\ Size})\\ &+ \gamma_{012}\,(\mathit{Class\ Size}{*}\mathit{TRIADFT})\\ &+ \gamma_{013}\,(\mathit{Class\ Size}{*}\mathit{TRIADNFT}) + \mu\end{aligned}$$

where

β_{0j} is the mean outcome status adjusted for other predictors for children in school j;
γ_{00} is the intercept associated with level-1 predictors across schools;
γ_{01} is the slope associated with the REMA pretest across schools;
γ_{02} is the main effect for the Blocking variable across schools;

γ_{03} is the main effect of the TRIAD-FT condition across schools;
γ_{04} is the main effect of the TRIAD-NFT condition across schools;
γ_{05} is the main effect of the percentage of children receiving free and/or reduced lunch (FRL%) across schools;
γ_{06} is the interaction of FRL% and the TRIAD-FT condition across schools;
γ_{07} is the interaction of FRL% and the TRIAD-NFT condition across schools;
γ_{08} is the main effect of Site across schools;
γ_{09} is the interaction of Site and the TRIAD-FT condition across schools;
γ_{010} is the interaction of Site and the TRIAD-NFT condition in schools;
γ_{011} is the main effect of Site across schools;
γ_{012} is the interaction of Site and the TRIAD-FT condition across schools;
γ_{013} is the interaction of Site and the TRIAD-NFT condition in schools;
and μ is the random effect at level-2.

and

$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}\text{TRIADFT} + \gamma_{22}\text{TRIADNFT}$$
$$\beta_{2j} = \gamma_{30} + \gamma_{31}\text{TRIADFT} + \gamma_{32}\text{TRIADNFT}$$
$$\beta_{4j} = \gamma_{40} + \gamma_{41}\text{TRIADFT} + \gamma_{42}\text{TRIADNFT}$$

where

γ_{10} is the average effect for the REMA pretest across schools
γ_{20} is the average effect for Gender (Male) across schools
γ_{21} is the cross-level interaction term between Gender and TRIAD-FT
γ_{22} is the cross-level interaction term between Gender and TRIAD-NFT
γ_{30} is the average effect for Ethnicity (AA) across schools
γ_{31} is the cross-level interaction term between Ethnicity (AA) and TRIAD-FT
γ_{32} is the cross-level interaction term between Ethnicity (AA) and TRIAD-NFT
γ_{40} is the average effect for having an IEP across schools
γ_{41} is the cross-level interaction term between IEP and TRIAD-FT
γ_{42} is the cross-level interaction term between IEP and TRIAD-NFT