

Longitudinal Evaluation of a Scale-Up Model for Teaching Mathematics With Trajectories and Technologies: Persistence of Effects in the Third Year

Douglas H. Clements

Julie Sarama

University of Denver

Christopher B. Wolfe

Indiana University-Kokomo

Mary Elaine Spitler

University of Buffalo, SUNY

Using a cluster randomized trial design, we evaluated the persistence of effects of a research-based model for scaling up educational interventions. The model was implemented in 42 schools in two city districts serving low-resource communities, randomly assigned to three conditions. In pre-kindergarten, the two experimental interventions were identical, but one included follow-through in the kindergarten and first-grade years, including knowledge of the pre-K intervention and ways to build upon that knowledge using

DOUGLAS H. CLEMENTS is the Kennedy Endowed Chair in Early Childhood Learning and professor at the University of Denver Morgridge College of Education, Katherine A. Ruffatto Hall 154, 1999 East Evans Avenue, University of Denver, Denver, CO 80208-1700; e-mail: Douglas.Clements@du.edu. His research interests include the learning and teaching of early mathematics; computer applications; creating, using, and evaluating research-based curricula; and taking interventions to scale.

JULIE SARAMA is the Kennedy Endowed Chair in Innovative Learning Technologies and professor at the University of Denver Morgridge College of Education. Her research interests include young children's development of mathematical concepts and competencies, implementation and scale-up of educational reform, professional development models and their influence on student learning, and implementation and effects of software.

CHRISTOPHER B. WOLFE is an assistant professor of education psychology at Indiana University. His research interests include the development of reading and mathematics, curricula effects, and multilevel modeling.

MARY ELAINE SPITLER is a senior research scientist in the Early Math Research Lab at the University at Buffalo. Her research interests include young children's mathematical development and learning mechanisms in early childhood.

learning trajectories. Students in the experimental group scored significantly higher than control students ($g = .51$ for those who received follow-through intervention in kindergarten and first grade; $g = .28$ for non-follow-through), and follow-through students scored significantly higher than non-follow-through students ($g = .24$).

KEYWORDS: mathematics, early childhood, primary, at-risk students, classroom research, scale-up, learning trajectories, technology, computers and learning

Education needs generalizable models of scale-up and longitudinal research evaluating the persistence of the effect of their implementations (Borman, 2007; Cuban & Usdan, 2003; McDonald, Keesler, Kauffman, & Schneider, 2006), especially in early mathematics (National Mathematics Advisory Panel, 2008). We created a research-based model for scale-up called TRIAD (Technology-enhanced, Research-based, Instruction, Assessment, and professional Development; see Sarama, Clements, Starkey, Klein, & Wakeley, 2008; Sarama, Clements, Wolfe, & Spitler, 2012), with the intent to generalize to both other subject matter areas and other grade levels. We implemented in the early childhood years because long-range benefits to children are greatest for interventions in that period (Carneiro & Heckman, 2003; Clements & Sarama, 2009) and because there is conflicting evidence as to whether the effects of early interventions persist or fade. The present study is an experimental evaluation of this implementation, the third and final year of TRIAD's follow-through intervention.

The rationale for situating the intervention in mathematics is that children from low-resource communities and who are members of linguistic and ethnic minority groups demonstrate significantly lower levels of mathematics achievement than children from higher-resource, nonminority communities (Denton & West, 2002; Duncan & Magnuson, in press; National Mathematics Advisory Panel, 2008). Such differences are evident from the earliest years (National Research Council, 2001, 2009; Sarama & Clements, 2009). This is important because early mathematics competence predicts later achievement even into high school (National Mathematics Advisory Panel, 2008; National Research Council, 2009; Stevenson & Newman, 1986), with persistent difficulties with mathematics one of the strongest predictors of failure to graduate high school and enter college (Duncan & Magnuson, 2011). Interventions to address these early differences may benefit low-resource and minority children more because they have fewer educational opportunities in their homes and communities (Brooks-Gunn, 2003; Carneiro & Heckman, 2003; Natriello, McDill, & Pallas, 1990; Raudenbush, 2009). Such interventions have been shown to be effective; however, most have not been taken to scale and often used the individual child as the unit of analysis, despite their assignment to treatments by class or school, which can inflate findings (Case, Griffin, &

Kelly, 1999; for a review, see Clements & Sarama, 2011; Griffin & Case, 1997; Griffin, Case, & Siegler, 1994; Klein, Starkey, Clements, Sarama, & Iyer, 2008; National Mathematics Advisory Panel, 2008; Starkey, Klein, & Wakeley, 2004). Thus, an important practical and theoretical challenge 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.

The TRIAD model was created to address this challenge. In this study, we evaluated the persistence of effects of a TRIAD implementation, with and without the follow-through component, in its third and final year. That is, two experimental groups received the TRIAD implementation in pre-kindergarten (pre-K), and one of these groups also received the follow-through intervention in their kindergarten and first-grade year. Here we report the effects at the end of the students' first-grade year and whether these effects were equivalent for various subpopulations, as well as possible indirect, or mediational, effects through particular pedagogical practices (Clements, Sarama, Spitler, Lange, & Wolfe, 2011).

Theoretical Framework

TRIAD's theoretical framework (Sarama et al., 2008) is an elaboration of the network of influences theory (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, through phases of introduction, initial adoption, implementation, and institutionalization.

The TRIAD model scales up the support of "interactions among teachers and children around educational material" (Ball & Cohen, 1999, p. 3). This strategy creates extensive opportunities for teachers to focus on mathematics, goals, and students' thinking and learning, which improves teachers' knowledge of subject matter, teaching, and learning and increases student achievement (Ball & Cohen, 1999; Cohen, 1996; Schoen, Cebulla, Finn, & Fi, 2003; Sowder, 2007). Research suggests that the most important feature of a high-quality educational environment is a knowledgeable and responsive adult and that professional development can foster these characteristics (Darling-Hammond, 1997; National Research Council, 2001; Sarama & DiBiase, 2004; Schoen et al., 2003; Sowder, 2007). Use of demonstrations, practice, and feedback, especially from coaches, increases the positive effects of information-only training (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; Pellegrino, 2007; Showers, Joyce, & Bennett, 1987). The professional development in TRIAD provides a promising path for developing teachers' understanding of learning, teaching, curriculum, and assessment by focusing on research-based models of students' thinking and learning (cf. Bredekamp, 2004; Carpenter & Franke, 2004; Hiebert, 1999; Klingner,

Ahwee, Pilonieta, & Menendez, 2003). Research-based learning trajectories are TRIAD's core (Clements & Sarama, 2004a, 2004b; Clements, Sarama, & DiBiase, 2003). Learning trajectories have three components: a goal (that is, an aspect of a mathematical domain students should learn), a developmental progression of levels of thinking, and instruction that helps them move along that trajectory. Thus, they facilitate teachers' learning about mathematics, how students think about and learn this 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 (Clements & Sarama, 2009; Sarama & Clements, 2009) and help teachers focus on the "conceptual storyline" of the curriculum, a critical element that is often missed (Heck, Weiss, Boyd, & Howard, 2002; Weiss, 2002).

The full TRIAD model (see Sarama et al., 2008) includes guidelines for promoting equity through the use of curriculum and instructional strategies that have demonstrated success with underrepresented populations (Kaser, Bourexis, Loucks-Horsley, & Raizen, 1999) and promoting communication among key groups around a shared vision (Fixsen, Blase, Naoom, & Wallace, 2009; Fullan, 1992; Hall & Hord, 2001; Huberman, 1992; Kaser et al., 1999; Snipes, Doolittle, & Herlihy, 2002). The model also requires maintaining frequent, repeated assessment ("checking up") efforts and follow-through efforts emphasizing the purpose and expectations of the project. That is, key groups are involved in continual improvement through cycles of data collection and problem solving (Fixsen et al., 2009; Fullan, 1992; Hall & Hord, 2001; Huberman, 1992; Kaser et al., 1999; Snipes et al., 2002).

The complete TRIAD model also addresses the issue of persistence of effects. Evidence regarding the persistence of early interventions is mixed, with some researchers reporting positive and long-lasting effects (Broberg, Wessels, Lamb, & Hwang, 1997; Garces, Currie, & Thomas, 2002; Gray, Ramsey, & Klaus, 1983; Magnuson & Waldfogel, 2005; Montie, Xiang, & Schweinhart, 2006) and others claiming that the effects fade in the primary grades (Currie & Thomas, 1995; Fish, 2003; Natriello et al., 1990; Preschool Curriculum Evaluation Research Consortium, 2008; Turner, Ritter, Robertson, & Featherston, 2006; U.S. Department of Health and Human Services—Administration for Children and Families, 2010). A recent meta-analysis on fadeout involving nearly 1,100 effect sizes taken from 65 studies reported that impacts are reduced by about .04 standard deviation units per year, which implies that program impacts persist for about 10 years (Leak et al., 2012). Some researchers explicitly or implicitly attribute such fadeout to the evanescence of the effects per se, perhaps due to their inadequate potency (Fish, 2003; Natriello et al., 1990; Turner et al., 2006; U.S. Department of Health and Human Services—Administration for Children

and Families, 2010). The TRIAD model is based on the hypothesis that such claims are conceptually flawed and inaccurate and that many present educational contexts (e.g., minimal demands of curricula, standards, and teaching practices) unintentionally undermine persistence of measured effects of early interventions (cf. Brooks-Gunn, 2003). For example, after a successful pre-K experience, children may experience kindergarten and first-grade classrooms in which both the teachers and curricula assume little mathematical competence and target only early-developing skills (even without such pre-K experience, kindergarten and first-grade instruction often covers material children already know; Carpenter & Moser, 1984; Engel, Claessens, & Finch, in press; Van den Heuvel-Panhuizen, 1996). Teachers may remain unaware that some of their students have already mastered the material they are about to “teach” them (Bennett, Desforjes, Cockburn, & Wilkinson, 1984; Clements & Sarama, 2009; National Research Council, 2009; Sarama & Clements, 2009; Thomas, 1982). Even if teachers are so aware, pressure to increase the number of students passing minimal competency assessments may lead some teachers to work mainly with the lowest performing students. Within this context and without continual, progressive support, early gains appear to fade. To ameliorate these potentially deleterious factors, TRIAD provided the kindergarten and first-grade teachers in one treatment group knowledge of the pre-K intervention and ways to build upon that knowledge using learning trajectories. This was not a full curricular and pedagogical intervention as we implemented in the pre-K TRIAD classrooms, but rather a test of our hypothesis that helping primary grade teachers build upon the pre-K work would support the persistence of effects. The next section briefly describes the previously reported evaluations of our implementation of TRIAD in students’ pre-K and kindergarten years.

Evaluations of the First 2 Years

The TRIAD implementation occurred in 42 schools in two urban districts serving low-income communities, randomly assigned to three conditions. Here we review the results from students’ pre-K year. We then discuss the issue of the persistence of effects and the rationale for the follow-through components and end with a review of the results from the kindergarten year.

Pre-K year. 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. Students in the TRIAD groups learned more mathematics than the control students (effect size, $g = .72$; see Clements et al., 2011). Of possible moderators (including school percentage of free/reduced lunch and limited English proficiency, gender, and ethnicity variables), only one was significant: African American students learned less than other students in control classrooms and more than other students 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 students, 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, Lange, Clements, & Wolfe, 2012).

Kindergarten year. As an initial test of the alternative hypotheses regarding persistence of effects, we evaluated the effectiveness of TRIAD's follow-through intervention. Kindergarten teachers were taught about what children learned previously and ways to build upon it. They used the districts' written curricula, but were taught learning trajectories, especially 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, 2007/2012), the same suite that students had used previously, which follows the learning trajectories through the primary grades. We first performed "intent-to-treat" (ITT) analyses, in which the condition group (school) to which the student was originally assigned was maintained as that student's condition, regardless of how many days the student experienced in that school. These analyses estimate the lower bound of potential effects of the intervention as the focus is on the impact of the random assignment, ignoring anomalies such as crossover of students between treatment conditions. ITT analyses showed that at the end of kindergarten, students in the follow-through condition, but not the non-follow-through condition, scored statistically significantly higher than students in the control condition ($g = .33$; for a full report, see Sarama, Clements, et al., 2012). We also performed "treatment-on-the-treated" (TOT) analyses that estimate the effects for students who experienced the full duration of the condition to which they were originally assigned. Both groups outperformed the control condition in TOT analyses ($g = .38$ for the follow-through, $g = .30$ for the non-follow-through). In these analyses, only one moderator was statistically significant, with African American students within the TRIAD follow-through group scoring significantly better on kindergarten outcomes than African American students in the TRIAD non-follow-through group. The intervention's effects at the end of kindergarten were mediated by the number of specific mathematical activities and the classrooms' mathematics culture.

Present Study—Analyses of Effects in First Grade

Given the longitudinal nature of the issues of fading effects, and especially given the substantial difference between mathematics standards and curricula in pre-K and kindergarten versus first grade (e.g., increased formality of the mathematical expectations in first grade) and the possibility that mathematical competencies developed in the pre-K intervention play an

even more important role in first grade than in kindergarten (e.g., pre-K learning of numerical and geometric composition and decomposition may particularly support the first-grade focus on addition and subtraction), it is important to evaluate the effects of the TRIAD follow-through intervention into children's first-grade year. The current analysis examines these effects, utilizing ITT and TOT analyses to investigate differences in mathematics achievement between follow-through, non-follow-through, and control conditions. Given our previous findings, we continue the examination of African American identification as a moderator for the impact of treatment group. Further, although multiple studies have investigated gender as moderator on mathematical competency, the overall results are mixed. Some have found slight advantages for male students within distinct mathematical skills (Royer, Tronsky, Jackson, & Horace Marchant, 1999), in use of abstract strategy use (Fennema, Carpenter, Franke, & Levi, 1998), or at higher grades (Martens, Hurks, Meijs, Wassenberg, & Jolles, 2011). As this analysis considers the sample at first grade and signals a change in the expectations and content associated with mathematics instruction in traditional elementary schools, the current analysis sought to capture any developing differences in mathematical competencies across gender.

At the school level, we include the percentage of children receiving free or reduced lunch. As we purposely sampled within low-resources communities, the majority of our schools had more than 80% of their children receiving assistance for lunch. Also, although individual data on language competency with English was not consistently collected across all schools, we include the school-level percentage of children identified with English as a second language or with limited English proficiency. Further, although 4.6% of children took the pre-K pretest in Spanish and 3.2% took the pre-K posttest in Spanish, none of the children continued to need a Spanish version of the mathematics assessment across the subsequent time points. Finally, we continued to evaluate the mediational effects of the number of specific mathematical activities and the classrooms' mathematics culture, both statistically significant in prekindergarten and kindergarten, as well as other possible impacts of early childhood classroom environments.

We addressed three primary research questions.

Research Question 1: What is the persistence of effects of the TRIAD intervention, with and without follow-through, on achievement at the end of first grade? Do students in the TRIAD follow-through (TRIAD-FT) group on the average outperform students in the TRIAD non-follow-through (TRIAD-NFT) group in mathematics achievement at the end of first grade? Do students in each of these experimental groups outperform those in the control group?

Research Question 2: Are there significant moderators of any statistically significant effects? Do the effects of the three conditions differ by gender or ethnic group?

Research Question 3: Do measures of the classroom environment mediate effects of different treatments on achievement? Do specific measures of the quantity and quality of the schools' mathematics environment and teaching mediate the effects of treatment group on mathematics achievement?

Method

We used a cluster randomized trial (at the school level) experimental design to test TRIAD's impact across the varied settings. We made a list of eligible schools (all those that had not participated in prior Building Blocks or TRIAD research or development projects) and recorded their earliest school-wide standardized mathematics scores (statewide assessments of fourth graders) as an indicator of possible differences in schools' emphasis on and achievement in mathematics. Schools were rank ordered on these scores separately within each site and blocks created to contain three schools with similar scores. We (publicly, 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 used hierarchical linear modeling (HLM), capitalizing on the nested nature of the data, to evaluate the effects of the interventions on students' mathematics performance trajectories, account for possible variations of the effects among subgroups, and assess the mediational role of the classrooms' environment and teaching (Raudenbush, 2007, 2008).

Participants and Contexts

Participants were the 1,305 students in 106 classrooms from the original randomly assigned 42 schools in two urban school districts in the Northeast United States (Clements et al., 2011) and the pre-K to first-grade teachers in those schools. These public pre-K classrooms were housed in the same school buildings as the kindergartens and first grades. Table 1 describes these diverse populations. By the end of first grade, 1,127 students from 347 classrooms in 172 schools were tested. Of these, the 1,079 who completed first grade and both components of the assessment were included in the ITT analysis. All 42 schools were represented in the ITT group of 1,079, with the three treatment groups maintaining their original percentages. We used this population for ITT analyses in which the condition group (school) to which the student was originally assigned was maintained as that student's condition, regardless of how many days the student experienced in that school. Of these students, 750 remained within their randomized school (and thus in treatment condition) from preschool through first grade. We used this subpopulation for TOT analyses. These 750 students also

Table 1
Demographics of Participating Schools

	Intent-to-Treat Sample				Treatment-on-Treated Sample							
	N	Female	AA	Other	Percentage FRL	Percentage LEP	N	Female	AA	Other	Percentage FRL	Percentage LEP
All	1,079	551 (51.10%)	586 (54.30%)	493 (45.70%)	84.39	11.02	750	399 (53.20%)	426 (56.80%)	324 (43.20%)	84.72	11.08
TRIAD-FT	383	203 (53.00%)	229 (59.80%)	154 (40.20%)	83.76	10.44	262	139 (53.10%)	154 (58.80%)	108 (41.20%)	82.60	10.96
TRIAD-NFT	375	188 (50.10%)	198 (52.80%)	177 (47.20%)	85.11	9.36	253	132 (52.20%)	148 (58.50%)	105 (41.50%)	86.58	8.89
Control	321	160 (49.80%)	159 (49.50%)	162 (50.50%)	84.43	13.65	235	128 (54.50%)	124 (52.80%)	111 (47.20%)	85.10	13.56

Note. AA = African American; percentage FRL = percentage of children receiving free or reduced lunch at the school level; percentage LEP = children identified as either an English language learner or as having limited English proficiency; TRIAD-FT = TRIAD follow-through; TRIAD-NFT = TRIAD non-follow-through.

represented all 42 original schools, again with the three research groups maintaining their original percentages.

Both districts used a revision of the *Investigations* (*Investigations in Number, Data, and Space*, 2008) mathematics curriculum. Both districts had policies and procedures that limited teachers' flexibility in implementing the *Investigations* curriculum. For example, both district mathematics departments wrote and disseminated "pacing guides" that dictated what should be taught each week.

The TRIAD Follow-Through Intervention—First-Grade Year

TRIAD staff provided professional development to 40 TRIAD-FT first-grade teachers from the 12 originally randomized follow-through schools for approximately 7 half-day sessions (a total of 32 hours, about 8 of which dealt with data collection and other logistics for the research). Teacher attendance at these sessions averaged 74% (mode, 83%; median, 80%). At the beginning of the academic year, first-grade teachers were informed that some of the students enrolled in their classrooms had experienced the Building Blocks prekindergarten curriculum and were now participating in a longitudinal study. These teachers were expected to teach their district-assigned curriculum, differentiating their teaching as needed to facilitate continued growth in mathematical learning.

We designed the professional development based on research, meeting about once per month, providing curriculum materials, and attempting to engender a sense of "personal satisfaction" (Sarama & Clements, 2009, pp. 348–349; Sarama & DiBiase, 2004). Pedagogical strategies included lecture, demonstration/modeling, observing children's mathematical thinking and learning, and classroom practice via video, games, and role play (cf. Wolfe, 1991). Professional development began with the mathematical content and developmental progressions for each major mathematical topic, including counting, adding, subtracting, subitizing, shape, length, and early algebra. Following whole group presentations, small groups of teachers focused on instructional tasks from their *Investigations* units. They presented case studies from their own classrooms, connecting evidence of children's thinking to the learning trajectories and determining instructional tasks that would enable students to learn the ideas and skills needed to achieve curricular goals. Staff mentors encouraged teachers to use formative assessment based on the learning trajectories. Through monthly visits, the mentors provided classroom-based support for teachers in making individual and classroom-wide software assignments and in differentiating their teaching. The technological component of the intervention featured accessibility to Building Blocks software activities for students, and, for teachers, an online Building Blocks Learning Trajectories (BBLT) Web application. BBLT provides scalable access to the learning trajectories by means of descriptions,

videos, and commentaries (Sarama, Clements, et al., 2012; see also TRIADScaleUp.org) and allows teachers to view the learning trajectories through a curriculum or developmental perspective.

In contrast, teachers in the other two conditions, the “business-as-usual” control and the TRIAD-NFT, as well as the teachers in other schools to which students transferred after pre-K, did not receive follow-through activities for first grade.

Measures

School mathematics environment and teaching. 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; for a review, see Kilday & Kinzie, 2009; Sarama et al., 2008). There are 28 items, all of which are 5-point Likert scales. Assessors blind to treatment group observed and evaluated the mathematics activities for approximately half the day. Subscales include classroom culture, specific mathematics activities (SMA), the overall number of SMAs observed, and the total time during the observation devoted to mathematics instruction. Assessors complete the classroom culture section once to reflect their entire observation. This section includes two subsections, the first of which includes the environment (showing signs of mathematics) and interactions (e.g., adults interacted with and were responsive to children). An example item in the second subsection, “personal attributes of the teacher,” is “The teacher appeared to be knowledgeable and confident about mathematics (i.e., demonstrated accurate knowledge of mathematical ideas and procedures, demonstrated knowledge of connections between, or sequences of, mathematical ideas).” (See Appendix A in the online journal for the Classroom Culture subscale.) Assessors complete an SMA form for each observed mathematics activity, defined as one conducted intentionally by the teacher involving several interactions with one or more students or set up or conducted intentionally to develop mathematics knowledge. An example item in this section is, “The teacher began by engaging and focusing children’s mathematical thinking (i.e., directed children’s attention to, or invited them to consider, a mathematical question, problem, or idea).” The number of SMAs and total time on mathematics are recorded once for each observation.

To protect against drift, project staff led experienced COEMET observers in reviewing the instrument and manual and practicing administering the COEMET on videotaped classroom vignettes, checking agreement on each item. 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 28-item instrument, with .74 for the Classroom Culture scale (considered acceptable, especially given the small number of items across diverse indicators; Cicchetti, 1994) and .90 for the SMA scale (Clements et al., 2011; Clements & Sarama, 2008a). Predictive validity was supported by a regression in which the Classroom Observation total score accounted for a significant amount of the variance in students' posttest achievement scores after accounting for pretest scores, $F(2, 1304) = 384.5$, $p < .001$, $r^2\Delta = .045$ (Clements & Sarama, 2008a).

Students' mathematical knowledge. The Research-based Early Maths Assessment (REMA) (Clements, Sarama, & Liu, 2008) measures core mathematical abilities of students from age 3 to 8 years using an individual interview format, with standardized administration protocol, videotaping, coding, and scoring procedures. Abilities are assessed according to theoretically and empirically based developmental progressions (National Research Council, 2007; Sarama & Clements, 2009). Topics 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. Geometry topics include shape recognition, shape composition and decomposition, congruence, construction of shapes, and spatial imagery, as well as geometric measurement, patterning, and reasoning. The developmental progression of items within each trajectory as well as the fit of individual items has been reported in earlier research (Clements et al., 2008). The REMA measures mathematical competence as a latent trait in item response theory (IRT), yielding a score that locates students 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). The 225 items are ordered by Rasch item difficulty; students stop after four consecutive errors on each of the number and geometry sections. Based on the expected growth in mathematical competency from pre-K to first grade, administration at this time point began with item 30 of the number section and item 6 in the geometry section. A basal rule of six in a row correct was employed for both sections.

Training sessions on the REMA included orientation, demonstration, and practice, with a focus on standardized delivery. Subsequent individual practice sessions were taped and critiqued, with 98% to 100% error-free delivery required for certification. All assessment sessions were videotaped and each item coded by a trained coder for correctness and for solution strategy; 10% of the assessments were double-coded. Both assessors and coders were blind to the group membership of the children. Continuous coder calibration by an expert coder (one tape per coder per week) militated against drift. Calibration feedback was sent to coders, alerting them to any variance from coding protocols. Previous analysis of the assessment data showed that its reliability ranged from .93 to .94 on the total test scores (Clements et al., 2008); on the present population, the reliability was .92. In addition,

the REMA had a correlation of .86 with a different measure of preschool mathematics achievement (Clements et al., 2008), the Child Math Assessment: Preschool Battery (Klein, Starkey, & Wakeley, 2000), and a correlation of .74 with the Woodcock-Johnson Applied Problems subscale for a pre-K-specific subset of 19 items (Weiland et al., 2012).

Procedures

This evaluation is of the final year of the project's TRIAD implementation (previous procedures and analyses reported in Clements et al., 2011; Sarama, Clements, et al., 2012; Sarama, Lange, et al., 2012). The same staff conducted the teacher training, classroom observations, and student assessments and thus had substantial experience. Further, all assessors were recertified on their respective instruments.

Mainly due to children transferring to other schools, the number of teachers increased from 106 in pre-K to 275 teachers in kindergarten and 347 in first grade; schools expanded from the original 42 to 140 in the kindergarten year and 172 in the first-grade year. To meet financial constraints, we conducted COEMET observations in 93 of these classrooms, overselecting classrooms with larger proportions of children, sampling randomly within blocks where not all classrooms were selected (see details in Sarama, Clements, et al., 2012). We blocked the classrooms by the number of study students in the classrooms and sampled disproportionately, dividing each of four blocks (>8 , 5–8, 2–4, or <2 students) into the three conditions: TRIAD-FT (41% of sample), TRIAD-NFT (31% of sample), and control (28% of sample). Original percentages of sample per research group were TRIAD-FT 36%, TRIAD-NFT 34%, and control 30%. Assessors administered the REMA to students in the spring.

Results

Research questions will be addressed in several sections. First, we examine the effects of the TRIAD interventions, compared to the control condition, across pre-K, kindergarten, and first grade. We present both ITT and TOT analyses. Demographic composition of each group is displayed in Table 1. Second, we measure just the effects of the TRIAD-FT treatment by comparing the TRIAD-FT and TRIAD-NFT interventions (covarying out pre-K posttest scores). In both these sections, we assess potential moderators. An indicator variable for block membership is also included in each analysis to control for the slight variation in the fourth-grade math scores of the included schools at the beginning of the study (Kirk, 1982). HLM analyses (Raudenbush, Bryk, Cheong, & Congdon, 2006) included the use of pre-K pretest REMA scores (see Clements et al., 2011) as covariates at both the child level and (mean aggregated) at the school level for the ITT and TOT models. (See Appendix B in the online journal for the HLM model.)

The pre-K covariate at the child level was group centered and all other predictors were grand mean centered (Enders & Tofighi, 2007). Effect sizes are given in Hedges g , which accounts for treatment groups of different sizes. It is calculated by dividing the individual predictor beta coefficient resulting from the HLM analyses by the pooled standard deviation of each outcome variable (Hedges & Hedberg, 2007). Cross-level interactions were examined utilizing online interaction utilities (Preacher, Curran, & Bauer, 2006).

Effects of TRIAD Over 3 Years

Intent-to-treat analyses: Main effects. ITT analyses assessed the effect of the TRIAD interventions across three school years, using the pre-K pretest mathematics achievement score as a covariate and the end of first grade mathematics achievement scores as the dependent variable. The unconditional HLM, in which no predictors are included, indicated that about 20% of the variance in first-grade scores lay between schools ($\sigma^2 = .386$, $\tau = .094$, $\rho = .197$). The addition of the group centered pretest at the child level and grand centered school aggregated pretest at the school level accounted for 30% more variance in first-grade outcome scores ($\sigma^2 = .291$, $\tau = .045$, $\rho = .134$). To control for variability in school demographics, the percentage of children receiving free/reduced lunch and percentage of children identified as having limited English proficiency (LEP) was included. A blocking variable identifying which randomized block each school belonged to also was included. Together, these control variables, representing the baseline model utilized in subsequent analyses, accounted for 6% of the total residual variance in child outcomes ($\sigma^2 = .291$, $\tau = .024$, $\rho = .075$).

In the process of refining our final model, multiple child-level moderators were tested. Of the multiple ethnic groups represented within our sample, children identifying as African American constitute over half of the overall ITT sample (54%). Further, the comparison of African American to an amalgamated “other” including Caucasian (19%), Hispanic (21%), Asian (4%), and Other (2%) was the only ethnic comparison to demonstrate a significant group difference across models. Gender was also conserved for analysis across models as an exploratory factor. The final highly parsimonious model therefore included a dichotomized variable for female or not and another variable for African American or not included at the child level. Further, interactions with the treatment groups were included in each model estimated as predictors at the school level on the child-level slopes. Treatment groups (e.g., TRIAD-FT, TRIAD-NFT, and control) were entered at the school level as dummy-coded indicators. As this research focuses on the comparison of three groups, a system with a single dummy code would not allow the examination of the between-treatment effects in conjunction with each treatment group as compared to the control group. Therefore, two parallel analyses were conducted on each of the following ITT and

TOT models. In the left hand panel (LP) of Table 2, to capture the comparison of the TRIAD-FT and control conditions relative to the TRIAD-NFT condition, two dummy-coded variables were utilized wherein in the first treatment variable, TRIAD-FT, was coded as 1 and control and TRIAD-NFT as 0 and in the second treatment variable, control was coded as 1 and TRIAD-FT and TRIAD-NFT as 0. TRIAD-NFT was chosen as the comparative conditions in this panel for ease of interpretation regarding our hypothesis of higher first-grade outcome scores for the TRIAD-FT condition and lower scores for the control condition. In the right hand panel (RP) of Table 2, the dummy coding system was modified to compare the control condition to both the TRIAD-FT and TRIAD-NFT conditions. In this model, the dummy code for the first treatment variable was coded as 1 for TRIAD-FT and as 0 for control and TRIAD-NFT. Similarly, for the second treatment variable, the TRIAD-NFT condition is coded as 1 and the control and TRIAD-FT conditions were coded as 0. This allowed us to examine our additional hypothesis that both treatment groups would demonstrate significantly higher outcome scores at the end of first grade than the control condition. Model fit tests for allowing random effects for slopes were not significant across test models; therefore, fixed effects from each subsequent model are discussed and displayed.

In the final ITT model (Table 2), both the pretest and pretest aggregate were significant predictors of child outcomes at the end of first grade. A small but significant impact for the school-level percentage of children receiving services as an English language learner was found ($\beta = .005$; $SE = .001$, $p = .001$). As this factor was not included in the original randomization and differences exist across conditions, its significance is not surprising. In addition, as this variable represents school-level percentage of children receiving services, the process by which this relates to overall child outcomes would require additional investigation. Still, in its current role as a control variable, its impact ($\beta = .005$) on children outcome scores is quite small relative to other covariates (i.e., pretest scores $\beta = .634$). Neither the percentage of children receiving free/reduced lunch nor the blocking identifier accounted for significant variance in mathematics scores.

A significant difference was found for each of the treatment group comparisons. Children within the TRIAD-FT condition significantly outperformed both the TRIAD-NFT (LP: $\beta = .124$; $SE = .056$, $p = .03$, $g = .18$) and control (RP: $\beta = .264$; $SE = .057$, $p < .001$, $g = .38$) conditions within the ITT sample on the mathematics outcome at the end of first grade. The TRIAD-NFT condition also demonstrated significantly higher mathematics scores relative to the control condition (RP: $\beta = .139$; $SE = .054$, $p = .01$, $g = .21$). Overall, the positive impact of the TRIAD-FT condition relative to the control condition is nearly two times larger than the statistically significant but smaller impact of TRIAD-NFT condition relative to control. Thus, the TRIAD-FT condition, as in previous research, continued to demonstrate

Table 2
Intent-to-Treat (ITT; N = 1,079) Hierarchical Linear Modeling Final Fixed Effects Model Outcomes and
Variance Components for Research-based Elementary Math Assessment (REMA) Mathematics First-Grade
Posttest Scores, With Pre-K Scores as Covariates at Both Levels

ITT	TRIAD-FT and Control Compared to TRIAD-NFT			TRIAD-FT and TRIAD-NFT Compared to Control		
	Coefficient	SE	p	Coefficient	SE	p
Intercept	-.012	.023	.615	-.012	.023	.615
Level 1 (child)						
Pretest	.382**	.021	.000	.382**	.021	.000
Gender (female)	-.032	.033	.342	-.032	.033	.342
Gender × TRIAD-FT	.097	.078	.211	.146	.081	.071
Gender × Control	-.049	.081	.659	.049	.081	.659
African American	-.252**	.040	.000	-.252**	.040	.000
AA × TRIAD-FT	.085	.092	.355	.191*	.094	.043
AA × Control	-.106	.090	.240	.106	.090	.240
Level 2 (school)						
Pretest	.634**	.098	.000	.634**	.098	.000
Aggregate						
TRIAD-FT	.124*	.056	.032	.264**	.057	.000
Control	-.139*	.054	.014	.139*	.054	.014
LEP	.005	.001	.001	.005	.001	.000
F/RL	-.004	.002	.090	-.004	.002	.081
Block	.005	.011	.659	.005	.013	.709
Variance Component	Random Effect	SD	p			
Level 1	.281	.530				
Level 2	.009	.095	.000			

Note. SE = standard error; SD = standard deviation; coefficient = unstandardized beta; TRIAD-FT = Building Blocks in pre-K with follow-through in kindergarten and first grade; TRIAD-NFT = Building Blocks in pre-K only; AA = African American; LEP = percentage of children identified as English language learner or with limited English proficiency; F/RL = percentage of children receiving free/reduced lunch; block = randomization block identifier based on fourth-grade math scores.

p* < .05; *p* < .01.

Table 3
Means and Standard Deviations of Mathematics Outcome
Scores by Treatment Condition and Time Point

Condition by Time Point					
TRIAD-FT	N = 262	TRIAD-NFT	N = 253	Control	N = 235
Pre-K pretest	37.64 (5.56)	Pre-K pretest	37.32 (6.02)	Pre-K pretest	37.96 (5.38)
Pre-K posttest	47.60 (4.35)	Pre-K posttest	47.21 (4.36)	Pre-K posttest	44.39 (4.71)
Kindergarten	53.42 (4.79)	Kindergarten	52.69 (4.18)	Kindergarten	51.76 (4.52)
First grade	60.72 (4.88)	First grade	59.32 (4.46)	First grade	59.04 (4.55)

Note. Rasch scores converted to a *T*-score ($M = 50$, $SD = 10$). TRIAD-FT = TRIAD follow-through; TRIAD-NFT = TRIAD non-follow-through.

a significantly larger impact on children outcome scores than the TRIAD-NFT condition.

Gender was not found to be a significant moderator within any of these comparisons. That is, within the ITT sample, treatment was found to be equally effective for children identifying with either reported gender. A main effect for African American was found to be significant (LP/RP: $\beta = -.252$; $SE = .040$, $p < .001$, $g = .39$). Overall, children identifying as African American performed significantly worse on the mathematics outcome as compared to children identifying with other ethnic groups. A significant difference, however, for children identifying as African American within the TRIAD-FT condition relative to control was found (RP: $\beta = .191$; $SE = .094$, $p = .043$). Children identifying as African American within the TRIAD-FT condition scored significantly higher than children identifying as African American within the control condition (relative to children identifying with another ethnic group), yielding a moderate effect size ($g = .31$). Overall, random assignment to the TRIAD-FT condition significantly reduces the negative main effect for African Americans found across conditions.

Treatment-on-the-treated analyses. All other analyses were conducted on the TOT sample; that is, they were performed on those students who stayed within their randomized condition and school, completed all components of the assessments, and did not repeat or skip kindergarten ($N = 750$). Table 3 presents the means and standard deviations of mathematics (REMA) scores by treatment condition and time point. An unconditional model based on the TOT sample indicated that 27% of the variance lay across schools ($\sigma^2 = .346$, $\tau = .126$, $\rho = .267$). An additional 25% of the variance was accounted for by the addition of the group centered pretest at the child level and school aggregated grand centered pretest at the school level ($\sigma^2 = .273$, $\tau = .079$, $\rho = .225$). Finally, the addition of the control variables (e.g., blocking identifier, school level percentage of free/reduced lunch, and school

level percentage of English language learners) accounted for an additional 9% of the total residual variance ($\sigma^2 = .273$, $\tau = .049$, $\rho = .154$).

The same child-level predictors of gender (i.e., female or not) as well as African American (i.e., African American or not) were examined within the TOT sample at the child level. This decision was based upon these two comparisons representing the largest subgroups within the sample. Within the TOT sample, females constituted 53% and children identifying as African American constituted 57% of the first graders remaining within their randomized school. Lastly, treatment indicators were again entered as dummy-coded variables representing the three treatment groups. The same comparison of dummy-coded treatment variables previously described for Table 2 is depicted in Table 4. In the left hand panel, the TRIAD-FT and control conditions are compared to the TRIAD-NFT condition. In the right hand panel the TRIAD-FT and TRIAD-NFT conditions are compared to the control condition. Final TOT model parameter estimates for both two-group comparisons are displayed in Table 4.

Again, both the child-level and school-level pretest aggregate scores were significant predictors at the end of first grade. Of the other control variables, the only significant main effect was found for the percentage of children receiving services for English language learning ($\beta = .006$; $SE = .002$, $p = .004$).

As in the ITT analyses, each of the treatment groups was significantly different from one another within the TOT sample. Relative to both the TRIAD-NFT (LP: $\beta = .163$; $SE = .081$, $p = .046$, $g = .24$) and control (RP: $\beta = .346$; $SE = .080$, $p < .001$, $g = .51$) conditions, children within the TRIAD-FT treatment group demonstrated the greatest mathematics competency at the end of first grade. Children within the TRIAD-NFT condition also demonstrated higher scores than the control group at the end of first grade (RP: $\beta = .183$; $SE = .080$, $p = .028$, $g = .28$). Within this sample, the impact of the TRIAD-FT condition relative to control was nearly twice that of the TRIAD-NFT condition relative to control. Attenuating the impact of the TRIAD-FT condition relative to the TRIAD-NFT condition by focusing on only those children who received the full treatment yields a one-third increase in effect as compared to the ITT estimates.

Gender did not act as a significant moderator in either set of comparisons, again suggesting that both boys and girls benefit similarly from exposure to the TRIAD-FT intervention. A significant moderate main effect for African American compared to other ethnic groups was found across both comparison sets (LP/RP: $\beta = -.240$; $SE = .049$, $p < .001$, $g = .38$). Across research groups, students identifying as African American scored lower on the assessment of mathematics achievement at the end of first grade controlling for pretest scores as compared to students identifying with other racial/ethnic groups. A significant interaction with the TRIAD-FT and control group comparison, however, was found for students identifying as African American compared to students identifying with other ethnic backgrounds (RP: $\beta = .290$; $SE = .120$, $p = .017$, $g = .38$). This moderate effect size indicates that children

Table 4
Treatment on the Treated (TOT; N = 750) Final Fixed Effects Model Outcomes and Variance Components for Research-based Elementary Math Assessment (REMA) Mathematics First Grade Posttest Scores, With Pre-K Scores as Covariates at Both Levels Including Each Treatment Group Comparison

TOT	TRIAD-FT and Control Compared to TRIAD-NFT			TRIAD-FT and TRIAD-NFT Compared to Control		
	Coefficient	SE	p	Coefficient	SE	p
Intercept	-.003	.033	.386	-.003	.033	.386
Level 1 (child)						
Pretest	.347**	.025	.000	.347**	.025	.000
Gender (female)	-.028	.039	.464	-.028	.039	.464
Gender × TRIAD-FT	.168	.092	.069	.149	.094	.114
Gender × Control	.019	.095	.845	-.019	.095	.845
African American	-.240**	.049	.000	-.240**	.049	.000
AA × TRIAD-FT	.061	.119	.609	.290*	.120	.017
AA × Control	-.187	.112	.095	.259*	.114	.023
Level 2 (school)						
Pretest aggregate	.664**	.136	.000	.664**	.136	.000
TRIAD-FT	.163*	.081	.046	.346**	.080	.000
Control	-.183*	.080	.028	.183*	.080	.028
LEP	.006	.002	.004	.006	.002	.004
F/RL	-.005	.003	.113	-.005	.003	.113
Block	-.001	.015	.949	-.001	.015	.949
Variance Component	Random Effect	SD	p			
Level 1	.262	.512				
Level 2	.026	.161	.000			

Note. SE = standard error; SD = standard deviation; coefficient = unstandardized beta; TRIAD-FT = Building Blocks in pre-K with follow-through in kindergarten and first grade; TRIAD-NFT = Building Blocks in pre-K only; block = randomization block identifier based on fourth-grade math scores; LEP = percentage of children identified as English language learner or with limited English proficiency; F/RL=percentage of children receiving free/reduced lunch; AA = African American.

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

identifying as African American differentially benefited from follow-through relative to children identifying as African American within the control condition. Focusing on only those students who have received the full treatment, inclusion within the TRIAD-FT condition eliminates the negative overall main effect on first-grade outcome scores for children identifying as African American. The results for these comparisons are displayed in Table 4.

Effects of (Only) the TRIAD Follow-Through Component

As the TRIAD-FT and TRIAD-NFT groups received the same intervention during the pre-K year, we calculated a third two-level model that included the pre-K posttest as a covariate at both the child and school levels to determine the isolated effect of the additional training received by the kindergarten and first-grade teachers within the TRIAD-FT condition, compared to the pre-K only experience of the TRIAD-NFT group (see Table 5). Beginning with the same model including pretest covariates only ($\sigma^2 = .273$, $\tau = .079$, $\rho = .225$), the addition of the end of pre-K posttest at both levels of the analysis accounted for a substantial amount of variance, an additional 57% of the variance in child outcome scores ($\sigma^2 = .182$, $\tau = .046$, $\rho = .200$). Finally, the same control variables representing the percentage of free/reduced lunch (F/RL), percentage of children receiving services for English language learning, and the blocking identifier were entered at the school level ($\sigma^2 = .183$, $\tau = .032$, $\rho = .151$), accounting for an additional 6% in the total residual variance. Thus, the posttest at the end of pre-K is a stronger predictor of end of first-grade math scores than the child's pre-intervention exposure level of mathematics. Further, this time point reflects one full year of treatment exposure for both the TRIAD-FT and TRIAD-NFT groups (i.e., the pre-K year). The final model included the same child-level moderators (i.e., gender and African American) as well as the dummy-coded treatment indicators. In this model, the TRIAD-FT and TRIAD-NFT conditions are compared to one another. To maintain inclusion of all schools and conditions, the TRIAD-FT variable was dummy coded as 1 and TRIAD-NFT and control as 0. The control variable was coded as 1 for control and 0 for TRIAD-NFT and TRIAD-FT conditions. Included together, the compared condition for this analysis is the TRIAD-NFT condition.

A significant difference between the TRIAD-FT and TRIAD-NFT conditions, controlling for both pretest and posttest, was found ($\beta = .169$; $SE = .075$, $p = .032$, $g = .25$). The TRIAD-FT condition demonstrated significantly higher outcome scores at the end of first grade relative to the TRIAD-NFT condition. This moderate effect size is comparable to the impact of the TRIAD-NFT condition relative to the control condition across the ITT and TOT models. Providing continuing professional development focused on mathematics learning trajectories to kindergarten and first-grade teachers helped children maintain the benefits of the pre-K intervention.

Table 5

“Value Added” Treatment on the Treated (TOT; $N = 750$) Final Fixed Effects Model Outcomes and Variance Components for Research-based Elementary Math Assessment (REMA) Mathematics First-Grade Posttest Scores, With Pretest and Posttest Pre-K Scores as Covariates at Both Levels Comparing TRIAD-FT to TRIAD-NFT

	Coefficient	<i>SE</i>	<i>p</i>
Intercept	-.023	.030	.444
Level 1 (child)			
Pretest	.089**	.025	.001
Posttest	.614**	.033	.000
Gender (female)	-.044	.032	.166
Gender \times TRIAD-FT	.145	.076	.056
Gender \times Control	-.007	.078	.925
African American	-.167**	.042	.000
African American \times TRIAD-FT	.068	.101	.504
African American \times Control	.014	.096	.882
Level 2 (school)			
Pretest aggregate	.403**	.154	.013
Posttest aggregate	.521**	.176	.006
TRIAD-FT	.169*	.075	.032
Control	.040	.105	.702
LEP	.005*	.002	.037
F/RL	-.002	.003	.509
Block	-.008	.014	.580
Variance Component	Random Effect	<i>SD</i>	<i>p</i>
Level 1	.178	.422	
Level 2	.025	.159	.000

Note. *SE* = standard error; *SD* = standard deviation; coefficient = unstandardized beta; TRIAD-FT = Building Blocks in pre-K with follow-through in kindergarten and first grade; TRIAD-NFT = Building Blocks in pre-K only; block = randomization block identifier based on fourth-grade math scores; LEP = percentage of children identified as English language learner or with limited English proficiency; F/RL = percentage of children receiving free/reduced lunch; AA = African American.

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

Of the control variables, only the percentage of children receiving services as an English language learner demonstrated a significant main effect at the school level ($\beta = .005$; $SE = .002$, $p = .037$). Although gender failed to demonstrate a significant main effect, the interaction between gender and treatment group, comparing TRIAD-FT to TRIAD-NFT, approached significance ($\beta = .145$; $SE = .076$, $p = .056$). Females within the TRIAD-FT condition

outperformed the males within both the TRIAD-NFT and control conditions ($g = .21$, Table 5).

A significant main effect was found for African American ($\beta = -.167$; $SE = .042$, $p < .001$, $g = .26$), whereby students who identify as African American demonstrated lower mathematics achievement scores at the end of first grade controlling for both pretest and posttest as compared to students identifying as a member of another ethnic/racial grouping. The interaction with the TRIAD-FT condition, however, was not significant ($\beta = .068$; $SE = .101$, $p = .504$). Therefore, when focusing on growth between kindergarten and first grade, there was no evidence that the TRIAD-FT treatment differentially supported African American students.

Mediation

We hypothesized that aspects of the classroom environment, as measured on the COEMET, serve to mediate the impact of the treatment condition on student outcomes, specifically, that the intervention (i.e., TRIAD-NFT and TRIAD-FT) operates through aspects of the classroom environment on student outcome mathematics scores collected at the end of first grade. Our analysis focused on three aspects of the COEMET that have previously been found to exhibit significant indirect effects on student outcomes on the same mathematics measure at the end of prekindergarten and kindergarten: the number of specific math activities, the quality of the specific math activities, and the classroom culture score. Again, to control for prior experience and improve our treatment effect outcomes, the pretest scores were covaried at both the student and school level in the mediational analysis. For ease of interpretation, to relate to our earlier findings, and to provide an upper limit estimate on possible indirect effects for classroom-level variables, these results present only pretest covariates. Further, the current two-level analysis of student nested within school necessarily obscures individual teacher differences within schools that may account for and allow for the investigation of other process level variables. Additional covariates would reduce the effect for each of the following models and represents a limitation in this analysis. These estimates therefore represent an upper bound on the impact of the measured school-level variables on child outcome scores at the end of first grade.

Utilizing notation (m) and (y) to reference the mediator and the outcome (Pituch, Stapleton, & Kang, 2006), the standard regression equation for the impact of the treatment group on the mediator can be

$$M_j = \gamma_{00(m)} + aX_j + u_{0j(m)},$$

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 $\gamma_{00(m)}$ and $u_{0j(m)}$, respectively. To estimate the impact of the mediator on child achievement posttest at the end of first grade, 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 first grade 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

$$\beta_{0j(y)} = \gamma_{00(y)} + \gamma_{01(y)} + c'X_j + bM_j + u_{0j(y)}$$

$$\beta_{1j(y)} = \gamma_{10(y)},$$

where $\gamma_{00(y)}$ is the intercept associated with the Level 1 pretest predictor and $\gamma_{01(y)}$ represents the slope associated with the REMA pretest across schools. 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 coefficients. The total effect of the independent variable, in this case treatment group, on the outcomes, first-grade math scores, is the sum of the indirect and direct effects. To estimate the impact of the mediation, the proportion of the overall effect accounted for by the indirect effect was also estimated (Fairchild, MacKinnon, Taborga, & Taylor, 2009; Pituch, Murphy, & Tate, 2010; Preacher & Hayes, 2008). Finally, we calculated 95% confidence intervals (CI) 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. This procedure has been found to demonstrate both high power and low Type 1 error rates (MacKinnon, Lockwood, & Williams, 2004). Table 6 presents the results.

We hypothesized that the TRIAD-FT condition would demonstrate greater indirect effects through these aspects of the classroom environment as this group of students continued in classrooms in which the teachers were instructed in the intervention. Increasing the quantity and quality of mathematics within the classroom environment is a primary focus of the intervention. Our results confirmed that the classroom components did not demonstrate any significant indirect effects for the TRIAD-NFT condition relative to control. The TRIAD-FT condition, however, did have significant indirect effects. A significant 20% of the total effect of the TRIAD-FT condition, relative to TRIAD-NFT, on student outcome scores was transmitted by scores on the classroom

Table 6

The *a* and *b* Path Coefficients for Kindergarten Classroom Observation of Early Mathematics Environment and Teaching (COEMET) Mediation Models, Including Significance Level of Each Path and 95% Confidence Intervals (CI) for the Indirect Effects

Comparison	<i>a</i>	<i>SE</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>ab</i>	95% CI
Classroom culture								
TRIAD-FT versus TRIAD-NFT	3.21	1.28	.017*	0.05	0.02	.005*	0.15	[0.02, 0.33]
TRIAD-FT versus control	3.18	1.21	.012*	0.04	0.02	.019*	0.12	[0.01, 0.27]
TRIAD-NFT versus control	0.27	1.29	.836	0.05	0.01	.002*	0.01	[-0.12, 0.15]
Number of SMAs								
TRIAD-FT versus TRIAD-NFT	1.09	0.98	.275	0.04	0.02	.051	0.04	[-0.04, 0.16]
TRIAD-FT versus control	1.90	0.90	.040*	0.03	0.02	.125	0.06	[-0.02, 0.18]
TRIAD-NFT versus control	0.88	0.92	.342	0.04	0.02	.035*	0.04	[-0.04, 0.15]
Quality of SMAs								
TRIAD-FT versus TRIAD-NFT	2.30	2.67	.394	0.00	0.01	.852	0.00	[-0.05, 0.07]
TRIAD-FT versus control	0.94	2.56	.714	0.00	0.01	.793	0.00	[-0.04, 0.05]
TRIAD-NFT versus control	-1.09	2.50	.667	0.00	0.01	.485	0.00	[-0.06, 0.04]

Note. Significant indirect effects are marked in bold. TRIAD-FT = TRIAD follow-through; TRIAD-NFT = TRIAD non-follow-through; SMAs = specific math activities.

**p* < .05.

culture from the student's experience in kindergarten (TRIAD-NFT indirect effect, .148; CI [.021, .328]). Similarly, as compared to the control condition, 35% of the total effect of the TRIAD-FT condition on outcome scores was transmitted through the classroom culture (indirect effect: .118; CI [.011, .274]). Exposure to the intervention during kindergarten served to improve the mathematical culture of the classroom, thereby increasing child outcome scores at the end of first grade. The other two components of the COEMET tested within this model, quality and quantity of specific math activities, did not demonstrate significant indirect effects. Relative to the control condition, however, the TRIAD-FT condition demonstrated significantly more specific math activities, though the indirect effect was not significant. None of the indirect effects examined for these COEMET components measured at first grade were found to be significant. The estimates for the *a* and *b* paths comprising each indirect effect for the kindergarten COEMET mediation analyses and their 95% confidence intervals are presented in Table 6.

Discussion

We need generalizable models of scale-up and longitudinal research evaluating the persistence of the effect of their implementations. This study was a cluster randomized trial evaluation of the persistence of effects of

a research-based model for scaling up educational interventions, with and without a follow-through intervention, into its third year of implementation. The instantiation of the TRIAD (Technology-enhanced, Research-based Instruction, Assessment, and professional Development) model was designed to teach early mathematics for understanding emphasizing learning trajectories and technological tools. The interventions were identical at pre-K, but only one included TRIAD's follow-through component in students' kindergarten and first-grade years.

We classified results into four categories. First, we discuss the effects of the interventions on mathematics achievement at the end of first grade, with intent-to-treat analyses that included all participating students, regardless of how much of the intervention they experienced. Second, we discuss the same effects with treatment-on-the-treated analyses that included only those students who experienced the assigned treatments. Third, we report whether these effects differed for different groups by examining several possible moderators of the effects. Fourth, we examined which, if any, measured components of the quantity and quality of the classroom environments and teaching mediated the effects of the TRIAD interventions.

First, addressing effects on students' mathematics achievement at the end of their first-grade year, the ITT analyses assessed the effect of the TRIAD interventions over the three school years, including all available students, regardless of whether they had switched classrooms, schools, or conditions. Results showed a significant difference between the TRIAD follow-through and the control group ($g = .38$) as well as the TRIAD-FT versus TRIAD non-follow-through ($g = .18$) comparison. The TRIAD-NFT and control conditions ($g = .21$) were also statistically significantly different. Even under the effects of dilution through mobility, the TRIAD intervention demonstrated significantly higher scores than the business as usual control condition and the follow-through component caused significantly higher scores than the non-follow-through condition.

Second, TOT analyses were conducted on all students who remained in their randomized condition throughout the 3 years. We contrast the ITT and TOT approaches, as the focus of this study is the effects of an intervention defined by consistently receiving mathematics instruction from educators engaged in professional development based upon learning trajectories of development within mathematics. Recall that at the end of pre-K, the TRIAD groups (TRIAD-FT and TRIAD-NFT, identical at pre-K) were statistically significantly higher in mathematics achievement than the control group (Clements et al., 2011). By the end of kindergarten, both groups similarly outperformed the control group. Our results at the end of first grade continue to support the benefits of high-quality, research-based mathematics instruction during pre-K as both groups continue to outperform the control conditions in TOT analyses. That is, with just the pre-K component, the TRIAD-NFT condition demonstrated significantly higher mathematics achievement than the

control condition ($g = .28$). Further, similar research-based instruction throughout the early years is more beneficial. At the end of first grade, the TRIAD-FT condition had significantly higher mathematics achievement than both the control condition ($g = .51$) and the TRIAD-NFT condition ($g = .24$).

Thus, extending the TRIAD intervention with follow-through to the end of first grade maintained the statistically significant gains with about the same effect sizes as measured at the end of kindergarten. In addition, these children scored significantly higher than those whose teachers implemented the pre-K but not the follow-through intervention. This evidence supports the persistence of the effects of this extended implementation of the TRIAD model, indicating that the follow-through component is important for maintaining the learning trajectory engendered by the pre-K intervention. Without the follow-through component, the effects are smaller each year.

Third, we examined potential moderators of gender and ethnic group (preliminary analyses confirmed previous results of no significant interaction with school percentage of free/reduced lunch). Multiple analyses revealed interactions between the TRIAD-FT intervention and African American students versus students identified with other ethnic groups. African American students within the TRIAD-FT group scored significantly better on student first-grade outcomes than African American students in the control group. Centering instruction around learning trajectories may focus teachers' attention on students' thinking and learning of mathematics rather than their memberships in ethnic groups and thus avoids perceptions that negatively affect teaching and learning (McLoyd, 1998; Pallas, Alexander, Entwisle, & Thompson, 1987; U.S. Department of Health and Human Services — Administration for Children and Families, 2010).

Examining only the effects of the follow-through intervention (children's kindergarten and first-grade years only, with pre-K posttest scores as a covariate), the TRIAD-FT condition exhibited greater mathematical competency than children within the TRIAD-NFT condition ($g = .25$). This indicates that the additional training provided for the kindergarten and first-grade teachers of these students continued to support children's development of mathematical competency. Overall, no interactions between ethnicity and the TRIAD-FT and TRIAD-NFT comparison were significant. That is, although all analyses that include the pre-K year indicate that African Americans score higher the more of the TRIAD intervention they experience, this interaction was not significant for analyses including only the kindergarten and first-grade years. This suggests that to continue to close the achievement gap between students who identify themselves as African American and other ethnic groups, additional work with primary grade teachers may be necessary, a point to which we will return. Finally, the appearance of a trend suggesting greater growth in mathematical competency for girls may warrant attention in future research. That is, given the change in mathematical expectations in first grade, documented differences in strategy use

(Fennema et al., 1998), the near significant finding for the follow-through component ($p = .056$), and the effect size of this interaction ($g = .21$), further investigation is needed.

Fourth, we examined what components of the classroom observation instrument mediated the effects. Only one statistically significant component of the instrument was found. A moderate portion of the total effect of the TRIAD follow-through intervention on student outcomes, relative to both the TRIAD-NFT and control conditions, was transmitted through the classroom culture observed in kindergarten. This is consistent with findings from the pre-K and kindergarten years. Exposure to the follow-through intervention may have created a greater focus on mathematics in these classrooms, which in turn increased student achievement. Such mediation is consistent with the literature indicating that learning is influenced by features of the classroom, especially overall mathematical activity and teachers who are not only knowledgeable and enthusiastic about mathematics, but also who interact with and respond to students' mathematical thinking (Clarke & Clarke, 2004; Clements & Sarama, 2007a, 2009; Fraivillig, Murphy, & Fuson, 1999; Sawada et al., 2002; S. P. Wright, Horn, & Sanders, 1997). Further, other components of the classroom observation that were significant mediators in previous work were not significant in this study. These results may be due simply to the limited number of classroom observations we could make across the large number of first-grade classes.

Implications

From the multiple perspectives of theory, research, policy, and practice, there is a need for transferable, empirically supported models of scaling up successful interventions (Borman, 2007; Cuban & Usdan, 2003; McDonald et al., 2006), particularly targeting young children's learning of mathematics (National Mathematics Advisory Panel, 2008). Multiple studies have supported the effectiveness of the TRIAD model (Clements et al., 2011; Sarama, Clements, et al., 2012; Sarama et al., 2008; Sarama, Lange, et al., 2012). This study provides the most compelling evidence to date regarding the importance of follow-through. With such follow-through, the effects from the pre-K intervention persisted; without follow-through, they were significantly smaller.

This finding has implications for the field. Multiple researchers have reported that preschool benefits do not persist; that is, that gains "fade" (Fish, 2003; Leak et al., 2012; Natriello et al., 1990; Preschool Curriculum Evaluation Research Consortium, 2008; Turner et al., 2006; U.S. Department of Health and Human Services—Administration for Children and Families, 2010)—a main rationale for the follow-through component and this study. We believe that such an interpretation mistakenly treats initial effects of interventions as independent of the students' future school contexts. That is, these researchers theoretically reify the treatment effect as an entity that should

persist unless it is “weak” and thus susceptible to fading. Such a perspective identifies the gain as a static object carried by the student that, if not evanescent, would continue to lift the student’s achievement above the norm. Our theoretical position and this study’s empirical results support an alternative view. Successful interventions do provide students with new concepts, skills, and dispositions that change the trajectory of the students’ educational course. However, these are, by definition, exceptions to the normal course for these students in their schools. Because the new trajectories are exceptions, multiple processes may vitiate their positive effects, such as institutionalization of programs that assume low levels of mathematical knowledge and focus on lower level skills and cultures of low expectations for certain groups (and, as noted, kindergarten and first-grade instruction often covers material children already know even without pre-K experience; Carpenter & Moser, 1984; Engel et al., in press; Van den Heuvel-Panhuizen, 1996). Left without continual, progressive support, children’s nascent learning trajectories revert to their original, limited course. An alternative explanation is that a stronger initial (pre-K) intervention is necessary to counteract any early disadvantage in requisite readiness skills. However, such an argument has been called unrealistic, especially if children attend poor-quality schools (Brooks-Gunn, 2003), which is more likely for African American students (Currie & Thomas, 2000). There is a cumulative positive effect of students experiencing consecutive years of high-quality teaching and a cumulative negative effect of low-quality teaching (Ballou, Sanders, & Wright, 2004; Jordan, Mendro, & Weerasinghe, 1997; Sanders & Horn, 1998; Sanders & Rivers, 1996; S. P. Wright et al., 1997). The latter is more probable for high-risk children (Akiba, LeTendre, & Scribner, 2007; Darling-Hammond, 2006). Further, the maintenance of the effect size in the TRIAD follow-through intervention, compared to the decreasing effect size in the non-follow-through intervention, militates against such an interpretation.

If supported by additional research, the finding has implications for both theory and policy. Interpretations of this fade may call for decreased funding and attention to pre-K (Fish, 2003, 2007). Although this may appear reasonable (e.g., “If effects of an intervention fade out, why fund that intervention?”), this ignores future school contexts. Instead, if such effects “fade” in traditional settings but do not in the context of follow-through interventions, then attention to and funding for interventions in both pre-K and the primary grades should arguably increase. This position is consistent with that of (a) the authors of the meta-analyses on fadeout, who conclude that because it takes a long time (about 10 years) for impacts to disappear, there is more than enough time for possible follow-through interventions that capitalize on the gains from these programs (Leak et al., 2012) and (b) intervention researchers’ notion of environmental maintenance of development (see also Cooper, Allen, Patall, & Dent, 2010; Ramey & Ramey, 1998). The follow-through component in this study cost about \$1,900 per teacher (this includes the costs of

the professional development staff, substitute teachers, and separately funded coaches for the year). More extensive and effective interventions are needed and may be achievable with similar funding.

In the evaluation of the same students in this study as well as previous studies, the TRIAD implementation was relatively more successful for students who identified themselves as African American than other ethnic groups. Although African American students continued to lag behind non-African American students in all conditions, the TRIAD-FT intervention helped them narrow that achievement gap. A high-quality, consistent mathematics education with an emphasis on learning trajectories can make a demonstrative and consistent positive impact on the educational attainment of African American students in the pre-K, kindergarten, and first-grade years compared to traditional instruction that is not accompanied by professional development based on student learning trajectories. We did not hypothesize this interaction, so we proffer explanations that are by necessity post hoc. (a) Centering instruction around learning trajectories may focus teachers' attention on students' thinking and learning of mathematics and what children can learn to do (Celedón-Pattichis, Musanti, & Marshall, 2010), avoiding biases, such as views of African American students' learning from a deficit perspective (Rist, 1970), that impair teaching and learning (Martin, 2007; McLoyd, 1998; Pallas et al., 1987; U.S. Department of Health and Human Services—Administration for Children, 2010). That is, especially given the significant mediation of the classroom culture, including enthusiastic interaction with children around mathematics they believe children can learn, it may be that the TRIAD interventions changed teachers' views of African American students' mathematical capabilities (Jackson, 2011). The curriculum's learning trajectories are based on the notion that learning is developmental and amenable to instruction, and the curriculum's approach, including specific, sequenced activities and formative assessment strategies, may have offered a way to act on these nascent views. In such action, the productive views are further strengthened. (b) The TRIAD intervention may promote a conceptual and problem-solving approach infrequently emphasized in schools serving low-income children (Stipek & Ryan, 1997), explicitly supporting African American students' participation in increasingly sophisticated forms of mathematical communication and argumentation (e.g., asking "How do you know?" as opposed to the "pedagogy of poverty" frequently used with African American students; Jackson & Wilson, 2012; Ladson-Billings, 1997). (c) The TRIAD follow-through intervention may raise several aspects of the quality of mathematics education, lack of which has been suggested as a reason preschool benefits dissipate for African American children (Currie & Thomas, 1995; see also Zhai, Raver, & Jones, 2012); for example, the language-rich nature of the curriculum and its expectation that all children invent solution strategies (cf. Carr, Steiner, Kyser, & Biddlecomb, 2008; Fennema et al., 1998) and explain them (such representations and

explanations promote future mathematics learning; Siegler, 1995). These and other possible reasons should be evaluated, compared, and combined, especially in interventions targeted to the primary grades.

The TRIAD follow-through intervention's effect was partially due to the increase in the positive classroom cultures teachers develop, at least in the kindergarten classrooms. Interventions such as TRIAD may 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. This is particularly important in the early years because teachers often do not recognize when tasks are too difficult, but even when they do, they provide “more of the same” (Bennett et al., 1984). Further, they overlook tasks that provide no challenge to children—that do not demand enough (Bennett et al., 1984; Van den Heuvel-Panhuizen, 1996). Thus, most children, especially those who have some number knowledge, may learn little or no mathematics in kindergarten (B. Wright, 1991).

Implementing interventions such as TRIAD is therefore important, given that early mastery of concepts and skills in mathematics and literacy is the best predictor of students' successful academic careers (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Duncan, Claessens, & Engel, 2004; Duncan & Magnuson, 2011; Paris, Morrison, & Miller, 2006; Stevenson & Newman, 1986). Further, students from low-income communities benefit more relative to students from higher resource communities from the same “dose” of school instruction (Raudenbush, 2009). Thus, comprehensive implementations of research-based models, such as the TRIAD follow-through model, may be especially effective in such lower resource schools. This speaks to a caveat concerning the effectiveness of the TRIAD follow-through intervention. The intervention maintained, but did not add to, the gains of the more comprehensive TRIAD pre-K intervention. Differences in scores remain statistically significant, but effects were not cumulative. Future design studies might investigate ways to (a) avoid or ameliorate the limiting influence of pacing guides and other school district policies that may have limited the effect of the TRIAD follow-through component, (b) increase the intensity or duration of that component, or (c) implement different and more extensive interventions, such as curriculum replacement (as the TRIAD intervention did in pre-K). That is, future research should evaluate the efficacy and scalability of a fully implemented TRIAD model in the primary grades to see if the pre-K slope can be maintained throughout elementary school. Such an intervention may go beyond “resisting fade out” to show that a positive rate of learning can and

should be sustained. In other words, we argue that what should persist is not just a pre-K gain, but also a dramatic trajectory of successful learning.

Findings for this and other implementations of the full TRIAD model, along with limitations of these studies, have five additional research implications. First, our research design could not identify which components of the TRIAD model and its instantiation are core components. Such research would be theoretically and practically useful. Second, in a related vein, the finding that TRIAD was particularly beneficial for African American children needs to be tested and, if replicated, explained. We proffered three possible reasons that might be tested in future research.

Third, this study addressed the persistence of TRIAD's effects. We also need studies of sustainability, the length of time an innovation continues to be implemented with fidelity (cf. Baker, 2007), especially given the "shallow roots" of many reforms (Cuban & Usdan, 2003). We presently are collecting such data.

Fourth, TRIAD was designed as a general model of scaling up successful interventions. TRIAD's 10 research-based guidelines are consistent with but more detailed than generalizations from the empirical corpus (Pellegrino, 2007). However, the model has not been implemented outside of the early childhood age range, or outside of urban districts with high percentages of students from low-resource communities, nor in other subject-matter domains. Evaluations of such varied implementations are needed. Fifth, the present study supports a guideline of the TRIAD model, the use of learning trajectories. That is, compared to the pre-K intervention's inclusion of a new curriculum with multiple components, the kindergarten and first-grade interventions included only one new curriculum component, the Building Blocks software. The core of the follow-through intervention was developing teachers' knowledge of the mathematical learning trajectories for their grade level. Thus, the TRIAD studies join the growing research corpus that supports the educational usefulness of learning trajectories, including evaluations of curricula built upon learning trajectories (Clements & Sarama, 2007b, 2008a; Sarama et al., 2008), elementary curricula based on related trajectories (e.g., Math Expressions in Agodini & Harris, 2010), studies of successful teaching (Wood & Frid, 2005), and professional development projects (Bright, Bowman, & Vacc, 1997; Clarke et al., 2002; R. J. Wright, Martland, Stafford, & Stanger, 2002). This supports the use of such structures in standards, such as the recently released *Common Core State Standards* (CCSSO/NGA, 2010). Again, however, the specific contribution of the learning trajectories per se needs to be disentangled and identified.

Note

This research was supported by the Institute of Education Sciences, U.S. Department of Education through Grant R305K05157. The opinions expressed are those of the authors

and do not represent views of the U.S. Department of Education. Although the research is concerned with the scale-up model, not particular curricula, a minor component of the intervention used in this research has been published by the authors, who thus could have a vested interest in the results. An external auditor oversaw the research design, data collection, and analysis, and other researchers independently confirmed findings and procedures. The authors wish to express appreciation to the school districts, teachers, and students who participated in this research and to Dr. Mark W. Lipsey for assistance with design issues and Dr. Larry V. Hedges for assistance with issues of statistical analyses.

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. doi:10.1080/19345741003770693
- Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher*, 36, 369–387. doi:10.3102/0013189X07308739
- 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. doi:10.1037/0022-0663.96.4.699
- 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.
- Ball, D. L., & Cohen, D. K. (1999). *Instruction, capacity, and improvement*. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.
- Ballou, D., Sanders, W. L., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1), 37–65. doi:10.3102/10769986029001037
- Bennett, N., Desforges, C., Cockburn, A., & Wilkinson, B. (1984). *The quality of pupil learning experiences*. Hillsdale, 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.
- Bredenkamp, 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. doi:10.1037/0012-1649.33.1.62
- Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report*, 17(1), 1, 3–14.

- Carneiro, P., & Heckman, J. J. (2003). Human capital policy. In A. B. Krueger & J. J. Heckman (Eds.), *Inequality in America: 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 T. K. Glennan, 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.
- Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15, 179–202.
- Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competence. *Learning and Individual Differences*, 18, 61–75.
- Case, R., Griffin, S., & Kelly, W. M. (1999). Socioeconomic gradients in mathematical ability and their responsiveness to intervention during early childhood. In D. P. Keating & C. Hertzman (Eds.), *Developmental health and the wealth of nations* (pp. 125–149). New York, NY: Guilford.
- CCSSO/NGA. (2010). *Common core state standards for mathematics*. Washington, DC: Council of Chief State School Officers and the National Governors Association Center for Best Practices. Retrieved from <http://corestandards.org/>
- Celedón-Pattichis, S., Musanti, S. I., & Marshall, M. E. (2010). Bilingual elementary teachers' reflections on using students' native language and culture to teach mathematics. In M. Q. Foote (Ed.), *Mathematics teaching & learning in K–12: Equity and professional development* (pp. 7–24). New York, NY: Palgrave Macmillan.
- Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.
- Clarke, D. M., Cheeseman, J., Gervasoni, A., Gronn, D., Horne, M., McDonough, A., . . . Rowley, G. (2002). *Early Numeracy Research Project final report*. Melbourne, Australia: 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 grades K–2: Painting a picture of challenging, supportive, and effective classrooms. In R. N. Rubenstein & G. W. Bright (Eds.), *Perspectives on the teaching of mathematics (66th yearbook)* (pp. 67–81). Reston, VA: National Council of Teachers of Mathematics.
- Clements, D. H., & Sarama, J. (2004a). Building Blocks for early childhood mathematics. *Early Childhood Research Quarterly*, 19, 181–189.
- Clements, D. H., & Sarama, J. (Eds.). (2004b). Hypothetical learning trajectories [Special issue]. *Mathematical Thinking and Learning*, 6(2).
- Clements, D. H., & Sarama, J. (2007a). 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, NY: Information Age Publishing.
- Clements, D. H., & Sarama, J. (2007b). 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, NY.
- Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York, NY: Routledge.
- Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. *Science*, 333(6045), 968–970. doi:10.1126/science.1204537
- Clements, D. H., & Sarama, J. (2012). Building Blocks software [Computer software]. Columbus, OH: SRA/McGraw-Hill. (Original work published 2007)
- Clements, D. H., Sarama, J., & DiBiase, A.-M. (2003). Preschool and kindergarten mathematics: A national conference. *Teaching Children Mathematics*, 8, 510–514.
- 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. doi:10.1080/01443410701777272
- 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.
- Cooper, H., Allen, A. B., Patall, E. A., & Dent, A. L. (2010). Effects of full-day kindergarten on academic achievement and social development. *Review of Educational Research*, 80(1), 34–70.
- Cuban, L., & Usdan, M. (Eds.). (2003). *Powerful reforms with shallow roots: Improving America's urban schools*. New York, NY: Teachers College.
- Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *American Economic Review*, 85, 341–364.
- Currie, J., & Thomas, D. (2000). School quality and the longer-term effects of Head Start. *Journal of Human Resources*, 35(4), 755–774.
- Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco, CA: Jossey-Bass.
- Darling-Hammond, L. (2006). Securing the right to learn: Policy and practice for powerful teaching and learning. *Educational Researcher*, 35(7), 13–24.
- 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
- Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In G. J. Duncan & K. Magnuson (Eds.), *Whither opportunity? Rising inequality and the uncertain life chances of low-income children* (pp. 47–70). New York, NY: Russell Sage Press.
- Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12, 121–138.

- Engel, M., Claessens, A., & Finch, M. (in press). Teaching students what they already know? The misalignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*.
- Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009). R^2 effect-size measure of mediation analysis. *Behavioral Research Methods*, 41, 486–498.
- Fennema, E. H., Carpenter, T. P., Franke, M. L., & Levi, L. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher*, 27, 6–11.
- Fish, R. (2003). *Effects of attending prekindergarten on academic achievement* (Unpublished master's thesis). University 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(5), 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 No. 231). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, the National Implementation Research Network.
- 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.
- Fullan, M. G. (1992). *Successful school improvement*. Philadelphia, PA: Open University Press.
- Garces, E., Currie, J., & Thomas, D. (2002). Longer term effects of Head Start. *American Economic Review*, 92(4), 999–1012.
- 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 bent . . . Lasting effects of preschool programs* (pp. 33–69). Mahwah, NJ: Erlbaum.
- Griffin, S., & Case, R. (1997). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education*, 3, 1–49.
- Griffin, S., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25–49). Cambridge, MA: MIT Press.
- Hall, G. E., & Hord, S. M. (2001). *Implementing change: Patterns, principles, and pot-holes*. Boston, MA: Allyn and Bacon.
- Heck, D. J., Weiss, I. R., Boyd, S., & Howard, M. (2002). *Lessons learned about planning and implementing statewide systemic initiatives in mathematics and science education*. Retrieved from http://www.horizon-research.com/presentations/2002/ssi_aera2002.pdf
- Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group randomized trials in education. *Educational Evaluation and Policy Analysis*, 1, 60–87.
- Hiebert, J. C. (1999). Relationships between research and the NCTM Standards. *Journal for Research in Mathematics Education*, 30, 3–19.
- 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.
- Jackson, K. (2011). *Exploring relationships between mathematics teachers' views of students' mathematical capabilities, visions of instruction, and instructional practices*. Paper presented at the American Educational Research Association, New Orleans, LA.

- Jackson, K., & Wilson, J. (2012). Supporting African American students' learning of mathematics: A problem of practice. *Urban Education*. Advance online publication. doi:10.1177/0042085911429083
- 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, H., Mendro, R., & Weerasinghe, D. (1997). *Teacher effects on longitudinal student achievement*. Paper presented at the National Evaluation Institute, Indianapolis, IN.
- Kaser, J. S., Bourexis, P. S., Loucks-Horsley, S., & Raizen, S. A. (1999). *Enhancing program quality in science and mathematics*. Thousand Oaks, CA: Corwin Press.
- Kilday, C. R., & Kinzie, M. B. (2009). An analysis of instruments that measure the quality of mathematics teaching in early childhood. *Early Childhood Education Journal*, 36, 365–372.
- Kirk, R. E. (1982). *Experimental design: Procedures for the social scientist*. Belmont, CA: Wadsworth.
- Klein, A., Starkey, P., Clements, D. H., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, 1, 155–178.
- Klein, A., Starkey, P., & Wakeley, A. (2000). *Child Math Assessment: Preschool Battery (CMA)*. Berkeley, CA: University of California, Berkeley.
- Klingner, J. K., Ahwee, S., Pilonieta, P., & Menendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional Children*, 69, 411–429.
- Ladson-Billings, G. (1997). It doesn't add up: African American students' mathematics achievement. *Journal for Research in Mathematics Education*, 28(6), 697–708.
- Lawless, K. A., & Pellegrino, J. W. (2007). Professional development in integrating technology into teaching and learning knowns, unknowns, and ways to pursue better questions and answers. *Review of Educational Research*, 77(4), 575–614.
- Leak, J., Duncan, G. J., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2012). *Is timing everything? How early childhood education program cognitive and achievement impacts vary by starting age, program duration and time since the end of the program*. Irvine, CA: University of California, Irvine Department of Education.
- 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(3), 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, 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.
- Martens, R., Hurks, P. P. M., Meijs, C., Wassenberg, R., & Jolles, J. (2011). Sex differences in arithmetical performance scores: Central tendency and variability. *Learning and Individual Differences*, 21, 549–554.
- Martin, D. B. (2007). Beyond missionaries or cannibals: Who should teach mathematics to African American children? *The High School Journal*, 91(1), 6–28.
- McDonald, S.-K., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). Scaling-up exemplary interventions. *Educational Researcher*, 35(3), 15–24.
- McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, 53, 185–204.

- 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.
- 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. (2007). *Taking science to school: Learning and teaching science in grades K–8*. 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.
- Pallas, A. M., Alexander, K. L., Entwisle, D. R., & Thompson, M. (1987). School performance, status relations, and the structure of sentiment: Bringing the teacher back in. *American Sociological Review*, 52, 665–682.
- 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., Murphy, D. L., & Tate, R. L. (2010). Three-level models for indirect effect in school and class randomized experiments in education. *The Journal of Experimental Education*, 78, 60–95.
- Pituch, K. A., Stapleton, L. M., & Kang, J. (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., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437–448.
- 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, 879–889.
- Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness (NCER 2008-2009)*. Washington, DC: Government Printing Office. Retrieved from <http://ncer.ed.gov>
- Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist*, 53, 109–120.
- 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(3), 169–180.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2006). *HLM: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

- Rist, R. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, 40(3), 411–451.
- Royer, J. M., Tronsky, L. N., Jackson, S. J., & Horace Marchant, I. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology*, 24, 181–266.
- Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247–256.
- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement* (Research Progress Report). Knoxville, TN: 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., Clements, D. H., Wolfe, C. B., & Spitler, M. E. (2012). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies. *Journal of Research on Educational Effectiveness*, 5(2), 105–135.
- Sarama, J., & DiBiase, A.-M. (2004). The professional development challenge in pre-school 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: Erlbaum.
- Sarama, J., Lange, A., Clements, D. H., & Wolfe, C. B. (2012). The impacts of an early mathematics curriculum on emerging literacy and language. *Early Childhood Research Quarterly*, 27, 489–502. doi:10.1016/j.jecresq.2011.12.002
- 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.
- Schoen, H. L., Cebulla, K. J., Finn, K. F., & Fi, C. (2003). Teacher variables that relate to student achievement when using a standards-based curriculum. *Journal for Research in Mathematics Education*, 34(3), 228–259.
- 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.
- Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology*, 28, 255–273. doi:10.1006/cogp.1995.1006
- 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.
- Sowder, J. T. (2007). The mathematical education and development of teachers. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 157–223). New York, NY: Information Age Publishing.
- Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, 19, 99–120.

- Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development*, 57, 646–659.
- Stipek, D. J., & Ryan, R. H. (1997). Economically disadvantaged preschoolers: Ready to learn but further to go. *Developmental Psychology*, 33, 711–723.
- Thomas, B. (1982). *An abstract of kindergarten teachers' elicitation and utilization of children's prior knowledge in the teaching of shape concepts*. Unpublished manuscript, School of Education, Health, Nursing, and Arts Professions, New York University.
- Turner, R. C., Ritter, G. W., Robertson, A. H., & Featherston, L. (2006, April). *Does the impact of preschool child care on cognition and behavior persist throughout the elementary years?* Paper presented at the American Educational Research Association, San Francisco, 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.
- Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*. Utrecht, the Netherlands: Freudenthal Institute, Utrecht University.
- Weiland, C., Wolfe, C. B., Hurwitz, M. D., Clements, D. H., Sarama, J. H., & Yoshikawa, H. (2012). Early mathematics assessment: Validation of the short form of a prekindergarten and kindergarten mathematics measure. *Educational Psychology*, 32(3), 311–333. doi:10.1080/01443410.2011.654190
- Weiss, I. R. (2002). *Systemic reform in mathematics education: What have we learned?* Paper presented at the meeting of the research presession of the 80th annual meeting of the National Council of Teachers of Mathematics Las Vegas, NV.
- Wolfe, B. (1991). *Effective practices in inservice education: An exploratory study of the perceptions of Head Start participants* (Unpublished doctoral dissertation). University of Wisconsin-Madison, WI.
- Wood, K., & Frid, S. (2005). Early childhood numeracy in a multiage setting. *Mathematics Education Research Journal*, 16(3), 80–99.
- Wright, B. (1991). What number knowledge is possessed by children beginning the kindergarten year of school? *Mathematics Education Research Journal*, 3(1), 1–16.
- Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press.
- Wright, R. J., Martland, J., Stafford, A. K., & Stanger, G. (2002). *Teaching number: Advancing children's skills and strategies*. London: 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.
- Zhai, F., Raver, C., & Jones, S. M. (2012). Academic performance of subsequent schools and impacts of early interventions: Evidence from a randomized controlled trial in Head Start settings. *Children and Youth Services Review*, 34(5), 946–954.

Manuscript received August 30, 2011

Final revision received September 21, 2012

Accepted October 17, 2012