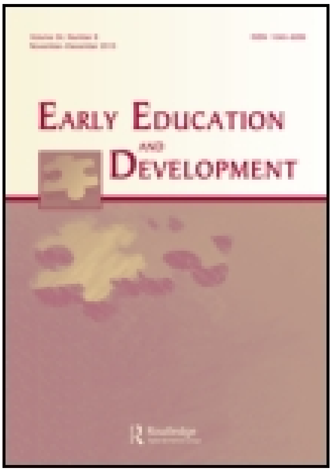

Early Education and Development
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/heed20

Sustainability of a Scale-Up Intervention in Early Mathematics: A Longitudinal Evaluation of Implementation Fidelity

Douglas H. Clements[a], Julie Sarama[a], Christopher B. Wolfe[b] & Mary Elaine Spitler[c]

[a] Educational Research, Practice and Policy, University of Denver

[b] Social Sciences Department, Psychology, Saint Leo University

[c] Learning and Instruction, University at Buffalo, The State University of New York

Published online: 30 Oct 2014.

Early Education and Development, 26: 427–449

ISSN: 1040-9289 print/1556-6935 online
DOI: 10.1080/10409289.2015.968242

Sustainability of a Scale-Up Intervention in Early Mathematics: A Longitudinal Evaluation of Implementation Fidelity

Douglas H. Clements and Julie Sarama
Educational Research, Practice and Policy, University of Denver

Christopher B. Wolfe
Social Sciences Department, Psychology, Saint Leo University

Mary Elaine Spitler
Learning and Instruction, University at Buffalo, The State University of New York

Research Findings: We evaluated the fidelity of implementation and the sustainability of effects of a research-based model for scaling up educational interventions. The model was implemented by 64 teachers in 26 schools in 2 distal city districts serving low-resource communities, with schools randomly assigned to condition. *Practice or Policy*: Although a logical expectation would be that, after the cessation of external support and professional development provided by the intervention, teachers would show a pattern of decreasing levels of fidelity, these teachers actually demonstrated increasing levels of fidelity, continuing to demonstrate high levels of sustained fidelity in their implementation of the underlying curriculum 2 years past exposure. Different profiles of variables predicted separate aspects of sustainability.

Scaling up successful educational interventions is a critical but challenging problem for practice, policy, and research. Evaluation of the fidelity of implementation of the core components of the intervention is essential, distinguishing between theory failure and implementation failure (Raudenbush, 2007). Longitudinal evaluation is particularly important for large-scale interventions because a full concept of scale requires not only consequential implementation and diffusion but also endurance over long periods of time and a transfer of responsibility from any external organization to the internal resources of a school district (Coburn, 2003; Dearing & Meyer, 2006). Such lasting effectiveness can be categorized as persistence or sustainability. We use *persistence* to mean the continuation of the effects of an intervention on individual children's trajectories of learning. We use *sustainability* to mean the continued use of program components in the achievement of desirable student outcomes and professional practice over time, with a focus on the maintenance of core beliefs and values, and the use of these core beliefs to guide

Correspondence regarding this article should be addressed to Douglas H. Clements, Morgridge College of Education, Marsico Institute for Early Learning and Literacy, University of Denver, Katherine A. Ruffatto Hall 154, 1999 East Evans Avenue, Denver, CO 80208-1700. E-mail: Douglas.Clements@du.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/heed.

adaptations (Century, Rudnick, & Freeman, 2012; Scheirer & Dearing, 2011; Timperley, 2011), with ongoing improvement in student outcomes over the length of time an innovation continues to be implemented with fidelity (cf. Baker, 2007). This is an especially important issue given the relationship between such fidelity and higher student outcomes (O'Donnell, 2008); the need for evaluations of sustainability past the implementation of the intervention, especially by developers or researchers (Baker, 2007; Glennan, Bodilly, Galegher, & Kerr, 2004; McDonald, Keesler, Kauffman, & Schneider, 2006); and the shallow roots of many reforms (Cuban & Usdan, 2003).

We created a research-based model for scale-up called TRIAD (Technology-enhanced, Research-based, Instruction, Assessment, and professional Development; see Sarama & Clements, 2013; Sarama, Clements, Starkey, Klein, & Wakeley, 2008; Sarama, Clements, Wolfe, & Spitler, 2012). Previous studies have documented positive effects of an implementation of the model on teachers' fidelity of implementation (Clements, Sarama, Spitler, Lange, & Wolfe, 2011) and on child outcomes (Clements et al., 2011; Sarama et al., 2008) as well as evidence that the child outcomes persist (Clements, Sarama, Wolfe, & Spitler, 2013; Sarama et al., 2012). The present study addressed two main questions. First, does the TRIAD model show sustained effects on implementation fidelity? Second, what factors significantly moderate the degree of sustainability?

BACKGROUND/CONTEXT

The theoretical framework supporting the scaling up of this research-based mathematics intervention is an interconnected network of influences at the school, classroom, and student levels (Sarama, Clements, Henry, & Swaminathan, 1996), with faithful implementation at the center of the model (Sarama et al., 2008). Sustained, faithful continuation of the intervention is theorized as a teacher outcome, measured as *fidelity of implementation*—''the determination of how well an innovation is implemented according to its original program design or as intended'' (Lee, Penfield, & Maerten-Rivera, 2009, p. 837). Learning trajectories are at the core of the components of the TRIAD intervention, guiding the written curriculum (*Building Blocks;* Clements & Sarama, 2007c, 2013), the instruction, the formative assessment, and the professional development. We define *learning trajectories* as

> descriptions of children's thinking and learning in a specific mathematical domain, and a related, conjectured route through a set of instructional tasks designed to engender those mental processes or actions hypothesized to move children through a developmental progression of levels of thinking, created with the intent of supporting children's achievement of specific goals in that mathematical domain. (Clements & Sarama, 2004b, p. 83)

Building on the work of hundreds of research studies (e.g., Carpenter, Franke, Jacobs, Fennema, & Empson, 1998; see Clements & Sarama, 2009; Sarama & Clements, 2009), the TRIAD/*Building Blocks* learning trajectories are not simply educated guesses but rather include empirically supported developmental progressions (more so for more heavily researched topics, of course). Researchers have theorized that coherence of research and connection to practice support sustainability (Boerst et al., 2010; O'Connell, 2012). Learning trajectory–based instruction provides coherence and offers an emerging theory of teaching (Sztajn, Confrey, Wilson, & Edgington, 2012). This emerging theory guided the analysis and interpretation of these sustainability data.

Research on Fidelity of Implementation

Theory and research on the fidelity of implementation has evolved from a notion of assumed fidelity (Dearing, 2008) to a more robust implementation science, affirming the growing recognition of the need to move from multiple definitions and approaches (Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012; O'Donnell, 2008) to a more unified, systematic evaluation framework, with compromised agreement on a common definition; standard, reliable, and valid measurements; and analyses of the links between fidelity and student outcomes (O'Donnell, 2008) that can account for the complexity of fidelity measurement (Burns, Peters, & Noell, 2008; Vartuli & Rohs, 2009) and the mediating and moderating effects of student (Raudenbush, 2007) and teacher (Achinstein & Ogawa, 2006) characteristics. Recently there has been an emphasis on the a priori identification of and subsequent measurement of fidelity to the unique and critical components of the intervention (Darrow, 2013; Nelson et al., 2012; O'Donnell, 2008; Shapley, Sheehan, Maloney, & Caranikas-Walker, 2010), with the understanding that evaluation theory is guided by program theory (Donaldson & Lipsey, 2006), which guides the determination of core components (O'Donnell, 2008).

Research on the Sustainability of Fidelity of Implementation

Educational researchers also are beginning to examine the sustainability of evidence-based programs, with the increasing understanding that scaling up exemplary programs may not be cost effective if these programs are not sustained (Fishman, Penuel, Hegedus, & Roschelle, 2011). There have been recent attempts to define and operationalize the construct of sustainability, including building generic conceptual models and theoretical frameworks with which to evaluate the sustainability of fidelity (Century et al., 2012; Fishman et al., 2011; O'Connell, 2012; Timperley, 2011) and to compare methods of the measurement of sustainability of fidelity (Sanford DeRousie & Bierman, 2012). Multiple synonyms for and definitions of sustainability have been identified, with increasing recognition of the need to construct a more scientific approach through consensus (Century et al., 2012; Scheirer & Dearing, 2011). In addition, there has been recent progress in the recognition of the ethical, economical, and beneficial outcome considerations in regard to sustainability, as well as the need for common ground in the interest of scientific progress on understanding sustainability across disciplines (Scheirer & Dearing, 2011). Progress in operationalizing the construct of sustainability is evident in recent efforts to identify the layers that influence sustainability: characteristics of the innovation itself, characteristics of the users, characteristics of the organization, and characteristics of the environment (Century et al., 2012; Scheirer & Dearing, 2011). There is growing recognition of the need not only to identify the core and customizable components of the innovation but also to measure the impact of the *sustained* implementation of these components on the outcomes and to identify the predictors of sustained fidelity.

In this study, given the nature of the measurements and data collection in our research design, we made an effort to analyze sustainability and the predictors of this sustained fidelity with a focus on program and user characteristics. Several user characteristics predictive of sustained implementation have been identified, including teacher perception of the benefits of the innovation (Fishman et al., 2011), teacher understanding of the core principles of the innovation and perceived coherence of the innovation (Century et al., 2012; Fishman et al., 2011; Klingner, Arguelles, Hughes, & Vaughn, 2001; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Scheirer, 2005; Sindelar, Shearer, Yendol-Hoppey, & Liebert, 2006), and teacher experience (Century et al., 2012; Gersten, Chard, & Baker, 2000; Rohrbach, Graham, & Hansen, 1993).

Our primary analysis in this study used a growth curve model to examine fidelity over the course of the intervention year and then, 2 years after program interactions and funding ceased, sustainability. To build this analytical model, we used fidelity/sustainability data collected at four time points, as well as theoretically based potential moderating influences, using information from data sets other than fidelity: teachers' perceptions of the benefits of the intervention, teachers' understanding of the core principles of the intervention, and teacher experience. The use of teacher-reported practices as a moderator representing evidence of teachers' understanding of the principles of the innovation enabled us to examine more closely the mechanisms behind the increasing fidelity and to better understand the variation in the sustained faithfulness to each component of the intervention. The examination of this theoretically and empirically identified predictor, as well as the other two (perceived benefits and teacher experience), permitted a more refined look at the impacts on sustained practices while maintaining the integrity of the growth curve models.

Teachers' perceptions that an innovation benefits students predict sustainability (Century et al., 2012; Fishman et al., 2011; Klingner et al., 2001; Owston, 2007). In one study, the predominant reason teachers gave for their continued use of the innovation was perceived benefits for the children (Sanford DeRousie & Bierman, 2012). A meta-analysis indicated that users' perceptions of benefits was one of the main factors influencing sustainability of implementation fidelity (Durlak & DuPre, 2008). In another meta-analytical study, perceived benefits were a positive factor in the sustainability of the intervention, even when, at times, evaluation did not show benefits (Scheirer, 2005). Klingner et al. (2001) found that the main reason given by teachers for sustaining their implementation of instructional practice 3 years after professional development and supports were withdrawn was their perception that students were benefitting from the sustained implementation. The larger study from which these data are drawn has documented significant positive change for children within the treatment condition relative to children within the control condition (Clements et al., 2011, 2013; Sarama et al., 2012; Simon, 1997). Our own *Revised Network of Influences* (Sarama et al., 2012) has been revised again as evidence comes in on the impact of the perceived benefits of the intervention on the sustained implementation of the program. The increased achievement of the students is noticed by teachers and administrators alike and creates its own sustainable buzz.

Teachers' understanding of the principles of the innovation also influences sustainability (Century et al., 2012). Research-based curricular programs coherently aligned with student development, professional development, standards, and district/school policies enhance understanding and may predict sustainability (Fishman et al., 2011). Teachers' increasing understanding of the principles of a coherent innovation and their growing ability to effectively implement the practices of the innovation play a role in its sustainability (Rohrbach et al., 1993). As teachers implement the practices of the innovation, implementation fidelity increases; implementation fidelity predicts sustained implementation fidelity (Bierman et al., 2013; Sanford DeRousie & Bierman, 2012). We have included teachers' self-reported professional development–driven teacher practices in our analyses as a representation of the enactment of their understanding of the innovation, reflecting this predictor of sustainability.

Teacher experience has been examined as a factor in the sustained faithful implementation of innovations, with mixed results. Some have found that experienced teachers are less likely to implement and sustain innovations than novice teachers (Gersten et al., 2000; Rohrbach et al., 1993; Vartuli & Rohs, 2009). Experienced teachers who have structured their knowledge and beliefs based on extensive prior teaching experiences were hesitant to restructure those beliefs and constructs to accommodate new principles and practices (van Driel, Beijaard, & Verloop, 2001). New and inexperienced teachers were stronger implementers and were more likely to continue use of a program than experienced teachers (Rohrbach et al., 1993). Huberman (1995) described the first 3 years of teaching as a survival stage when inexperienced teachers were more open to faithfully implementing an innovation in the hopes that it would serve to stabilize their instruction at a vulnerable time in their careers. Conversely, experienced teachers were skeptical of new curricula. Alternatively, Varol (2009) found that teachers with more teaching experience were more likely to implement the components of an innovation. Gersten et al. (2000) reported that experienced teachers between their seventh and 20th years of teaching have been found to be open to change due to a willingness to experiment during a time of stabilization, provided the change brings promise of greater effectiveness. Teacher inexperience was identified by teachers as a threat to the sustainability of an intervention (Bobis, 2006). However, experienced teachers may be ''entrenched in their ways'' (Bobis, 2006, p. 26), unwilling to incorporate the innovation into the routine and sustain it. These mixed results have led to our own inquiry as to the impact of teacher experience on the sustained faithful implementation of the *Building Blocks* program.

We place our current analysis within this context as we examine the faithfulness of teachers' implementation of the critical core components of the TRIAD program model over three time points during the study and teachers' implementation of these components at a fourth time point, 2 years after the end of the study (and thus the cessation of professional development and other external supports). The current study also addresses and examines three theoretical predictors of sustainability: teachers' perceptions of children's mathematics achievement; teachers' understandings of the principles of the innovation, as evidenced by the self-reported enactment of professional development–driven practices; and teacher experience.

RESEARCH QUESTIONS

Although there is growing interest in implementation fidelity, little research has been reported on the sustainability over time of faithful implementations of learning innovations (O'Connell, 2012). This study examined the sustainability of teachers' implementation fidelity to the core components of a prekindergarten mathematics intervention two full years after external support ceased. Some researchers have reported that fidelity decreases as time between interactions with developers increases (Datnow, 2005; Hargreaves, 2002). We hypothesized that an innovation based on coherent learning trajectories, wherein teachers increase their familiarity with mathematical goals, hypothesized developmental paths leading to the achievement of those goals, and activities designed to enable children to attain those goals, positively impacts teachers' sustained and faithful implementation of the innovation.

We ask two main questions. First, given that the TRIAD model generated good fidelity of implementation during the implementation of the study (Clements et al., 2011), does this implementation also demonstrate sustained effects, measured by observed fidelity of implementation 2 years after the research-based intervention was completed? That is, does fidelity decrease, remain consistent, or increase after the external supports provided by the initial research study have been completed and withdrawn? Second, what factors significantly moderate the degree of sustainability? Specifically, what role do the self-reported beliefs and practices of teachers within mathematics instruction play in sustaining fidelity? Does the perception of their children's mathematical growth support long-term fidelity to the intervention? Do differences in teaching experience impact persistent fidelity? Do these factors differ for different components of the curriculum that was implemented via the TRIAD model?

SETTING

This research draws from a large scale-up study of implementation fidelity and sustained fidelity across two diverse research settings involving teachers from 42 randomly assigned public schools (26 experimental) in two cities (one distal) serving primarily low-income, ethnically diverse children (see Clements et al., 2011, 2013; Sarama et al., 2012).

Participants

The original participants were from 106 (72 treatment) prekindergarten classrooms. At the time the final data set was collected, 2 years after the 72 treatment prekindergarten teachers participated in the study, 14 of the 72 had retired or were assigned to a different grade level; 164 was the maximum number of teachers with data by time point, by subsection ($M = 2.4$ per school). Data from 64 master's-level, all-but-one-female teachers in 26 schools across both sites are included here. About 89% had a master's degree or higher. They had an average of 17 years of teaching experience ($SD = 8$, range = 1–33), and their average number of years of prekindergarten teaching experience was 6.8. About 94% believed themselves to have the support of their principals. As a measure of morale, 76% of the teachers believed that ''staff members in this school generally have school spirit.'' The children they taught were 4-year-olds (51% female) of mixed ethnicity (53% African American, 21% Hispanic, 19% White, 3.7% Asian Pacific, 1.8% Native American, and 0.6% other). Most children (82.33%) received free or reduced lunch, 13.5% had limited English proficiency, and 10% had an individualized education plan. Treatment was assigned at the school level, randomly, by site. All prekindergarten teachers in each school participated. Each teacher implemented the intervention for all students in the class; however, data were collected on a maximum of 15 students per class (the average number of students participating in the study was 13.6 per classroom). District-mandated maximum class sizes differed by site (18 at Site 1, 24 at Site 2).

Intervention/Program/Practice

Teachers participated in 13 professional development days over 2 years (with about 2 days spent on research procedures rather than training) and interacted with project coaches as they implemented the curriculum. The focus in professional development sessions was not on just curricular instruction but on changing the way teachers thought about mathematics instruction by focusing on all three components of the learning trajectories: mathematical goal, developmental progression, and linked instruction (Clements & Sarama, 2004b, 2012; Clements et al., 2011).

Intervention curriculum. *Building Blocks* (Clements & Sarama, 2007a) is a National Science Foundation–funded mathematics curriculum designed using a comprehensive curriculum research framework (Clements, 2007) to address numeric/quantitative and geometric/spatial ideas and skills. Woven throughout areas are mathematical subthemes (such as sorting and sequencing) as well as mathematical processes (both general, such as communicating, reasoning, representing, problem solving, and overarching mathematizing; and specific, such as number or shape composition and patterning). These were determined to be critical mathematical building blocks (the same body of research and expertise guided the consonant Curriculum Focal Points; National Council of Teachers of Mathematics, 2006).

The *Building Blocks* instructional approach is finding the mathematics in, and developing mathematics from, children's activity (Clements & Sarama, 2007a). Children are guided to extend and mathematize their everyday activities, from block building to art to songs to puzzles, through sequenced, explicit activities (whole group; small group; centers, including a computer center; and throughout the day). Thus, off- and on-computer activities are designed based on children's experiences and interests, with an emphasis on supporting the development of mathematical activity. For more detail see Clements and Sarama (2004a, 2007a, 2007b).

The *Building Blocks* learning trajectories were designed to develop teachers' content knowledge by explicating the mathematical concepts, principles, and processes involved in each level and the relationships across levels and topics. For example, they introduce the components of geometric shapes (e.g., the correct definition of *side*) as well as relationships between components (e.g., sides forming a right angle) and shape classes (e.g., a square as a subcategory of rectangle and justification for this based on properties). The learning trajectories develop teachers' knowledge of children's developmental progressions in learning that content (moving from intuitively recognizing shapes as unanalyzed visual wholes, to recognizing components of shapes, to hierarchically classifying shape categories). They develop teachers' knowledge of the instructional activities designed to teach children the content and processes defining the level of thinking in those progressions and inform teachers of the rationale for the instructional design of each (e.g., why certain-length sticks are provided to children with the challenge to build specific shapes). The *Building Blocks* learning trajectories assist curriculum enactment with fidelity in that they connect the developmental progressions to the instructional tasks, providing multiple guidelines or sources of stability in teachers' instantiation of the instructional activities (cf. Ball & Cohen, 1996). Finally, the *Building Blocks* learning trajectories are designed to motivate and support the use of formative assessment.

Professional development. Although designed to support teachers' learning and implementation, the *Building Blocks* curriculum was not designed to be used in isolation from teacher training. Without training, teachers often fail to implement new approaches faithfully. For example, teachers often reduce the cognitive demand of instructional tasks after their initial introduction (Stein, Brover, & Hennigsen, 1996). As another example, teachers need training in understanding, administering, and especially using data from new assessment strategies, essential strategies in the effective use of learning trajectories for formative assessment (Foorman, Santi, & Berger, 2007).

In this project, teachers participated in 8 days of professional development during the school day in the first year (a gentle introduction to the curriculum with no child testing) and 5 days in the second year focused on the learning trajectories for each mathematical topic. Training addressed each of the three components of the learning trajectories. To understand the goals, teachers learned core mathematics concepts and procedures for each topic. For example, they studied the system of verbal counting based on cycling through 10 digits and the concept of place value (based on content similar to that presented in National Research Council, 2009; Wu, 2011). To understand the developmental progressions of levels of thinking, teachers studied multiple video segments illustrating each level and discussed the mental actions on objects that constitute the defining cognitive components of each level. To understand the instructional tasks, teachers studied the tasks and viewed, analyzed, and discussed video of these tasks as enacted in classrooms.

Each of these professional development contexts used the software application Building Blocks Learning Trajectories (BBLT), which presented and connected all components of the innovation. BBLT provided scalable access to the learning trajectories via descriptions, videos, and commentaries. The two sequential aspects of the learning trajectories, the developmental progressions of children's thinking and connected instructional tasks, are linked to each other. Discussions of BBLT classroom videos made explicit how such practice exemplified research-based principles.

This learning was sequenced according to the *Building Blocks* curriculum. Throughout this study, teachers learned how to use the learning trajectories as a basis for formative assessment, a key to high-quality teaching that has been shown to be particularly difficult for teachers to enact (Foorman et al., 2007). That is, they discussed and practiced how to interpret children's thinking and select appropriate instructional tasks for the class (e.g., compacting the curriculum if most children can learn it at a faster pace) and for individuals (e.g., assigning children to small groups or modifying activities within groups to match instructional tasks to the developmental levels of individual children).

In addition, project coaches observed and provided support to teachers and completed implementation fidelity evaluations. Coaches participated in the same professional development as the teachers. Before this, they participated in an additional day of professional development on coaching and administering the Fidelity instrument conducted by project staff. Additional meetings for coaches occurred throughout the year. Coaches then worked with teachers throughout the remainder of the project, visiting teachers in their classrooms about twice per month.

Research Design

Using longitudinal growth modeling, we examined fidelity to the curriculum at four time points across the course of the study: Fall 2006, Winter 2007, Spring 2007, and Spring 2009. The first three time points reflect observations of teacher fidelity during the study's data collection period. The Spring 2009 time point reflects a return to these prekindergarten classrooms 2 years past their involvement with the project's professional development and support.

Data Collection and Analysis

The main data source for this sustainability study was a classroom observation measure, rather than teacher self-reports, which made up the main body of evidence of sustainability (O'Connell, 2012; Timperley, Wilson, Barrar, & Fung, 2007). The Fidelity instrument was designed to evaluate whether, and to what degree, teachers were faithfully implementing the critical components of the curriculum, both on and off computer. The Fidelity instrument is composed of multiple subsections. Observers rate agreement with curriculum-based statements on a 5-point scale (0 = *neutral/not applicable*) for most items. In addition, each scale contains several dichotomous questions (1 = *present*, 0 = *not present*). The General Curriculum subscale includes five items for a possible total of 14 points ($\alpha = .60$–$.66$) focused on the overall use of the curriculum in the classroom, including compliance-oriented items (e.g., ''The teacher was within 2 weeks of the scheduled plan, based on the week they started, and adjusted for vacations/field trips/other days not available for mathematics'') and productive adaptations to the vision of the intervention (e.g., ''The teacher extended the activities in ways that enhanced the quality of the teaching and learning''). The Whole Group Activity subscale includes seven items for a total of 28 points ($\alpha = .81$–$.83$) focused on classroom-wide instruction (e.g., ''The teacher began by engaging and focusing children's mathematical thinking''). The Small Group subscale includes 25 items with a possible total of 75 points ($\alpha = .90$–$.95$) focused on mathematics instruction with three or four children (e.g., ''The teacher encouraged children to actively think, reason, solve problems, or reflect, as indicated in the written curriculum''). Finally, the Computer subscale includes 12 items for a possible total of 40 points ($\alpha = .73$–$.84$) focused on individual children's use of the Building Blocks software (e.g., ''The teacher monitored and/or observed children during the computer time'').

Project coaches collected data using this instrument after participating in several training meetings and conducting interrater reliability visits over 2 years. Interrater reliability was 95%, arrived at through simultaneous classroom visits by rotating pairs of observers for 10% of all observations. Four on-site classroom fidelity observations were conducted in the prekindergarten treatment classrooms using the Fidelity instrument. Three of these administrations occurred during the second intervention year (late fall, early winter, early spring). This provided the bulk of the fidelity evidence. The instrument was administered again in the same prekindergarten classrooms in the early spring, 2 years after the cessation of professional development and the withdrawal of funding. Observers were instructed to spend adequate time in the classroom to complete each section of the instrument (e.g., General Curriculum, Hands On Activities, Whole Group Activities, Small Group Activities, and Computer Activities) and to return to the classroom if necessary to cover any subsection not observed during the observation (e.g., computer activities that may have been offered only twice per week). These fidelity observations were conducted by the prekindergarten teachers' coaches.

Teachers' experiences, practices, beliefs, and knowledge. A teacher questionnaire measured teachers' self-reported knowledge, beliefs, and teaching practices pertaining to early childhood mathematics, including sections on demographics, teachers' education and experience, and school and teacher/classroom characteristics. The instructional practices included self-reports of frequency of use of instructional practices promoted in the professional development. For example, this included ''require students to explain their reasoning when giving answers,'' ''provide opportunities for students to discuss math with one another,'' ''provide opportunities for students to explore alternative methods for solutions,'' ''have students use multiple representations (e.g., numeric, graphic, geometric, etc.),'' and ''embed assessment in regular class activities.'' This scale included seven items measured on a 5-point Likert scale with reliability estimates ranging from .80 to .85. The data from this measure complemented the data collected onsite by the trained observer via the Fidelity instrument, providing an additional perspective and source of information regarding teachers' implementation of the innovation as they instantiated their understandings of the principles of the intervention.

Children's mathematical knowledge. The Research-based Elementary Math Assessment (Clements, Sarama, & Liu, 2008) measures the core mathematical abilities of students ages 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).

Previous analysis of the assessment data showed that their reliability ranged from .93 to .94 for the total test scores (Clements et al., 2008); in the present population, reliability was .92. In addition, the Research-based Elementary Math Assessment 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 prekindergarten-specific subset of 19 items (Weiland et al., 2012).

Children's gain in mathematics knowledge from pretest to posttest provides a proxy for teachers' perception of their children's learning. That is, teachers were blind to the study status of individual children within their classroom and were not told about individual children's mathematical growth or their classrooms' growth. Therefore, we define this parameter as teachers' own recognition of their students' learning of mathematics.

RESULTS

Model Creation and Specification

To assess fidelity to the original professional development program, we compared a series of latent growth models and growth mixture models for each of the Fidelity subscales. Three predictors were examined across these subscales. First, teacher experience was included as an averaged number of years of overall experience starting from the beginning of the study. Second, a sum score from the teacher questionnaire reflected teacher reports regarding instructional practices, with higher scores reflecting greater reported frequency of engaging in those practices promoted in the professional development. Finally, classroom-averaged child assessment scores from the end of prekindergarten were regressed on pretest scores to provide a metric of child mathematics achievement change across the time frame of the analysis.

An initial unconditional growth model was fit to the four time points to provide estimates of the average level of fidelity within each subscale and the average rate of growth over time for all teachers at the final time point of measurement (i.e., 2 years past direct interaction with the study; Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). A limitation of this type of analysis, however, is the assumption that all members of the sample are drawn from the same population and can be appropriately described by a single parameter for growth and level (Raudenbush & Bryk, 2002). To address this we implemented an additional growth mixture model analysis to capture differential relationships both within the measured treatment teachers and in response to the additional predictors of sustained fidelity. Coverage of relevant measures across the beginning, middle, and end of the intervention and 2 years after the intervention ranged from 89% to 60% within a given time point. Missing data were a result of resource limitations and were unrelated to the degree of fidelity or classroom operation. Thus, we classified these as missing at random, leading all models to be analyzed utilizing full maximum likelihood within Mplus 6.0 (L. K. Muthén & Muthén, 1998–2011).

Latent Growth Models

Table 1 contains the means and standard deviations for the subscales of the Fidelity measure across the four time points. The results for the basic growth curve analyses are displayed in Table 2. Given a focus on sustained fidelity 2 years past interaction with the program developers, time was centered at the final time point. Overall, average status on the subscales measuring fidelity to the different components of the intervention was significant at the final time point. For the General Curriculum subscale the average level of fidelity (13.39, $z = 18.41$) was 96% of the total possible score. Similarly, the Small Group (62.75, $z = 34.90$) and Computer (32.88, $z = 30.16$) subscales were 84% and 82% of the total possible within each scale. The rate of growth was also positive and significant for all of the subscales except for Whole Group. The Whole Group (18.64, $z = 46.92$) subscale demonstrated the lowest percentage of total fidelity possible for that subscale (66%) as well as a positive yet nonsignificant rate of growth. A quadratic acceleration factor was also tested for these basic models. This factor was significant only for the General Curriculum subscale (0.426, $z = 3.11$). Investigation of the raw and estimated means confirmed a significant jump in General Curriculum fidelity between the third and fourth time points. Alternatively, the addition of the quadratic acceleration factor did not demonstrate a good fit to the data for the other subscales and was eliminated. The quadratic factor was carried through for all subsequent analyses of the General Curriculum subscale.

TABLE 1
Number of Data Points, Means, and Standard Deviations for Subscales of the Fidelity Measure

	Time 1 (Pre-K Pre)		*Time 2 (Pre-K Mid)*		*Time 3 (Pre-K Post)*		*Time 4 (2 Years Post)*	
Subscale	*N*	M (SD)	*N*	M (SD)	*N*	M (SD)	*N*	M (SD)
General curriculum	61	10.31 (2.33)	49	9.43 (2.65)	64	10.50 (2.21)	53	13.39 (3.23)
Small group	49	45.76 (12.72)	45	44.24 (10.55)	53	51.72 (11.28)	48	62.65 (9.69)
Computer	43	23.72 (6.39)	32	25.19 (5.81)	45	26.67 (5.26)	41	32.56 (6.48)
Whole group	61	18.13 (2.16)	47	17.06 (2.75)	63	18.32 (1.75)	59	18.64 (2.34)

Note. *Data points* refers to the time points of data collection, designated Time 1, Time 2, Time 3, and Time 4. Pre-K = prekindergarten.

TABLE 2
Basic Growth Curve Results for Subscales of the Fidelity Measure

Coefficient	*General curriculum*	*Small group*	*Whole group*	*Computer*
Intercept	13.39**	62.75**	18.64**	32.88**
Slope	2.37**	5.99**	0.220	2.90*
Variance (intercept)	3.72**	8.40	2.47	21.74**
Variance (slope)			0.546*	
Quadratic	.426**	.004		−.053
Total possible	14	75	28	40
Percentage	96%	84%	66%	82%

Note. Negative variances were fixed to zero in the final presented models.
$^{*}p < .05$. $^{**}p < .001$.

Conditional growth model results incorporating the impact of child gain, teachers' stated positions on instructional practices, and years of experience teaching are displayed in Table 3. All predictors were mean-centered prior to entering the model. Child gain in mathematics achievement was a significant predictor of average status at the 2-year follow-up for all subscales of the Fidelity instrument. This suggests that teachers who perceived greater gain in mathematics within their classroom demonstrated higher levels of fidelity across the study time frame. Rate of growth was also positively and significantly impacted by child gain for both the General Curriculum and Computer subscales. A significant impact of child gain on the quadratic factor tested for the General Curriculum subscale indicated that teachers who perceived more growth in their children's mathematics achievement demonstrated steeper growth in General Curriculum fidelity. In these conditional models, teacher practices in mathematics instruction had a significant

TABLE 3
Conditional Growth Curve Results for Subscales of the Fidelity Measure With Effects for Years of Experience, Teacher Practices, and Child Achievement Gain

Coefficient	*General curriculum*	*Small group*	*Whole group*	*Computer*
Intercept	13.45**	63.29**	18.60**	32.65**
Slope	2.46**	6.06**	0.167	2.83**
Quadratic	.453**			
Variance (intercept)	9.96	28.49**	2.16	16.46
Variance (slope)	4.24		0.564*	1.06
Intercept on child gain	5.58**	10.35**	2.90**	9.17**
Slope on child gain	3.64**	−1.77	0.729†	1.69*
Quadratic on child gain	.777*			
Intercept on practices	0.263	0.504	0.091	0.623†
Slope on practices	0.152	−0.069	−0.006	0.278*
Quadratic on practices	.057			
Intercept on years experience	−0.004	0.215	−0.026	−0.022
Slope on years experience	0.020	0.172*	−0.001	−0.024
Quadratic on years experience	.004			

Note. Negative variances were fixed to zero in the final presented models.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .001$.

impact on the rate of growth only for the Computer subscale. That is, teachers highly endorsing frequency of professional development–driven practices demonstrated a higher rate of growth in fidelity to the computer component of the intervention across the time frame of the study. Finally, years of experience teaching had a significant impact on the rate of growth for the Small Group subscale. Teachers with more experience demonstrated a greater rate of growth in fidelity to the small-group components relative to teachers with fewer years of experience.

Growth Mixture Models

As mentioned earlier, a limitation of basic growth curve modeling is the assumption that all participants are drawn from a single group that can be characterized by a single status and rate of growth. In a longitudinal study of an intervention in which training is used to change the practices and beliefs of individuals, it is reasonable to assume that differing patterns in status and rates of growth may be obscured within a single group focus. Furthermore, these latent groups may demonstrate different relationships to the additional predictors of child gain, teacher practices, and years of experience. Based on the variability surrounding the growth factors (i.e., intercept and slope) within each subscale, all four were submitted for growth mixture modeling analysis (B. O. Muthén, 2004; Pickles & Croudace, 2010).

Two distinct groups[1] of teachers were discovered for the Fidelity subscales of General Curriculum, Small Group, and Computer. Three groups was determined to be the best fit for the Whole Group subscale. The number of groups was iteratively increased from the basic growth model of one utilizing differences in the Bayesian information criterion, whereby the *k*-1 model is preserved if the Bayesian information criterion is smaller than the *k* model. In addition, entropy values closer to 1 were evaluated as indicative of the presence of separable groups (Jung & Wickrama, 2008). Results from the growth mixture models can be found in Table 4.

The two groups identified for the General Curriculum, Small Group, and Computer subscales were characterized by their starting points at the beginning of prekindergarten (the first time point) as either low or high based on their initial status within each subscale. *High* refers to teachers with high levels of fidelity at the beginning of the study, whereas *low* refers to teachers with low fidelity at the beginning of the study. The trajectories for these subscales are plotted by class in Figure 1. Examining the predictor variables within each class for each subscale revealed a varied pattern. Within the General Curriculum subscale, a significant positive effect for child gain was found in both groups for status at the final time point (i.e., low = 42%, high = 58%). Teachers who perceived greater mathematics growth for their children had significantly higher levels of fidelity to the general curriculum than teachers who perceived less mathematics growth. Child gain was also significantly related to a higher rate of growth, but only for teachers within the high group. In addition, the high group demonstrated a significant impact for child gain on the quadratic factor, leading to acceleration in growth to the final time point. The only other significant impact of the predictors on the two groups within the General Curriculum subscale was for teacher practices. Teachers endorsing a higher frequency of behaviors such as encouraging children to describe their solution strategies (‘‘How do you know?’’), describe their results, and use

[1]We use the term *group* instead of the more usual statistical term *class* to avoid confusion with frequent references to the teachers’ class(rooms).

TABLE 4
Results From Two and Three Growth Mixture Models With Effects of Mean-Centered Child Mathematics Gain, Teacher Practices, and Years of Experience

	General curriculum		*Small group*		*Computer*		*Whole group*		
Coefficient	*Low*	*High*	*Low*	*High*	*Low*	*High*	*Low*	*Medium*	*High*
Intercept	10.95**	15.28**	57.33**	60.76**	31.96**	33.42**	17.26**	18.87**	16.94**
Slope	2.23*	2.86**	6.92**	5.15**	4.42**	2.58**	0.802	0.238*	−2.01**
Quadratic	.457	.496*							
Intercept on child gain	5.29**	4.52*	21.66*	9.55*	24.44*	7.75*	1.61	3.22**	14.20
Slope on child gain	2.90	3.50*	2.98	−1.19	3.41	2.55*	−0.366	0.951*	4.66
Quadratic on child gain	.400	.915*							
Intercept on practices	0.434*	0.200	−2.75**	2.29**	−5.66*	0.748*	−0.034	−0.038	−0.131
Slope on practices	−0.188	0.143	−0.119	0.261	−1.71	0.299*	−0.016	−0.011	0.014
Quadratic on practices	−.058	.082†							
Intercept on years experience	−0.023	0.200	0.318	0.131	−1.33**	0.043	0.125	−0.032	−0.575
Slope on years experience	0.030	0.734	0.612**	0.099*	−0.578**	0.623	0.075	−0.011	−0.238
Quadratic on years experience	.000	.694							
Final proportion	42%	58%	29%	71%	13%	87%	15%	75%	10%

†$p < .10$. *$p < .05$. **$p < .001$.

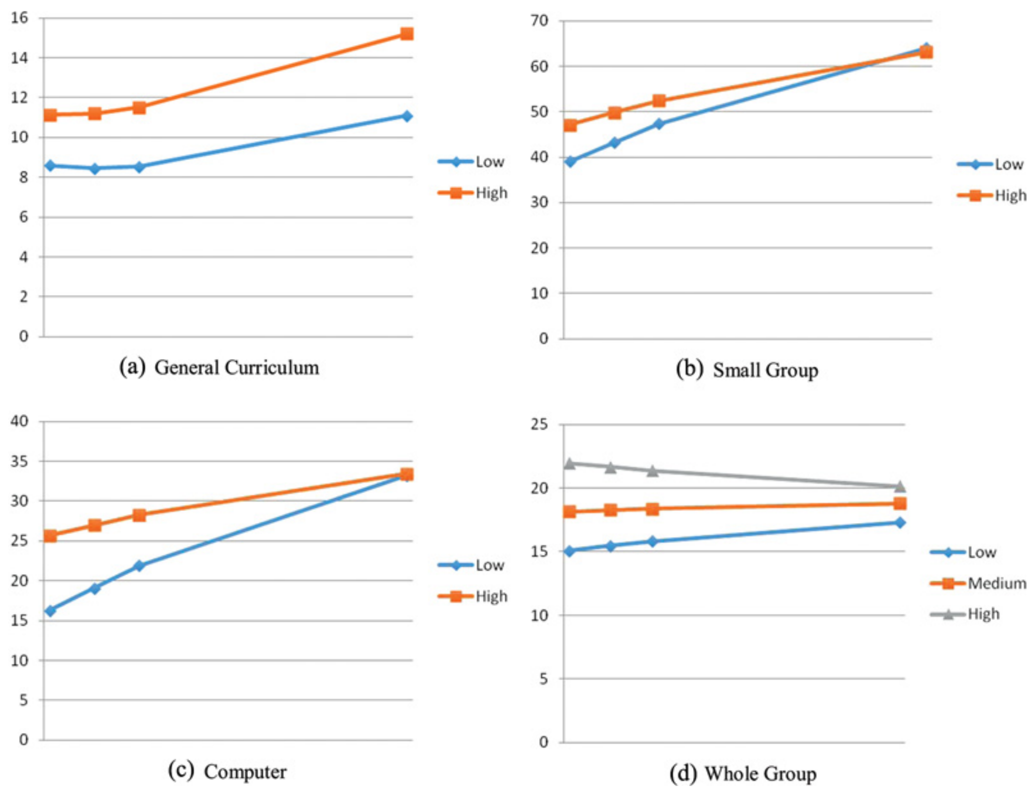

FIGURE 1 Subscale scores for classes, Times Points 1, 2, 3, and 4.

multiple solution strategies had a higher status in fidelity on this subscale 2 years after the intervention. Years of experience did not have a significant impact on intercept or slope within either group.

The two groups found within the Small Group (low = 29%, high = 71%) subscale also had a significant impact for child gain on final status 2 years after the intervention. The benefit was greater for teachers within the low group. The impact on the rate of growth for child gain was not significant. Teacher practices had a different impact on the growth factors of the two groups. Teachers within the low group who reported engaging in more professional development–driven behavior were lower in fidelity at the final time point than low-group teachers who did not report engaging in those behaviors. Conversely, teachers within the high group endorsing a higher frequency of instructional practices were significantly higher in Small Group fidelity. Finally, although neither group displayed a significant impact of years of experience on status 2 years past the end of the intervention, a significant positive impact on rate of growth was found for both groups. Teachers within the low group had a higher slope, however, suggesting greater growth in this group for teachers with more experience.

Again, average fidelity to the Computer subscale (low = 13%, high = 87%) was significantly and positively impacted by child gain. A larger impact for child gain on computer fidelity was found for the low teacher group. Overall, teachers perceiving greater mathematics growth demonstrated higher average fidelity to the components of the Computer subscale. Alternatively, teachers within the high group experienced a greater rate of growth when perceiving greater growth in their children's mathematics. Teachers within the low group reporting greater frequency of the instructional practices were lower in Computer subscale fidelity. Conversely, teachers within the high group evinced a benefit for higher frequency of reported use of these instructional practices on both final status and rate of growth. Finally, within the low group, teachers with more experience were lower on average on Computer fidelity as well as demonstrated a slower rate of growth from the beginning of prekindergarten to 2 years after the intervention.

Finally, fit statistics suggested three separate groups for the Whole Group subscale of the Fidelity measure. The three groups in this group were similarly characterized by their initial status in Whole Group fidelity (low = 15%, medium = 75%, high = 10%). Unlike the other subscales, in which both groups demonstrated positive growth, the high group for Whole Group had a significant negative rate of growth. Teachers within this group started high in fidelity but decreased in Whole Group fidelity. Teachers within the low group demonstrated a positive but not significant positive rate of growth. Both groups came to a central point in Whole Group fidelity similar to the significant, though slight, trajectory for the majority of teachers within the medium group. Child gain was a significant predictor for only the medium group, for which it had a positive impact on both average status and rate of growth. None of the other predictors had a significant impact within any of the groups for the Whole Group subscale.

DISCUSSION

Longitudinal evaluation of the fidelity of implementation in large projects is important because successful scale-up requires not only consequential implementation but also endurance over long periods of time (Coburn, 2003; Dearing & Meyer, 2006), without which the project may not be cost effective (Fishman et al., 2011). Therefore, our first research question was whether the implementation of the TRIAD model demonstrate sustained effects, measured by observed fidelity of implementation 2 years after the research-based intervention was completed. We found that teachers exposed to the TRIAD intervention continued to demonstrate high levels of fidelity to the underlying curriculum 2 years past the end of the external intervention, without continued project support.

The rate of growth across the 3-year period was also positive and significant for all of the subscales except for Whole Group. Growth appeared to be linear for the Small Group and Computer subscales, but General Curriculum fidelity showed accelerated growth between the third and fourth time points—that is, from the end of the intervention to the observation 2 years later. We posit that teachers first learned to implement the instructional components and made continual improvements in such implementation but made substantial growth in the General Curriculum subscale only when they had some skill and confidence in the separate instructional components. That is, they needed to have these pieces somewhat in place before they could synthesize them into a coherent, positive classroom culture for early mathematics.

Our second research question addressed factors that significantly moderated the degree of sustainability. Conditional growth models indicated that the average child gain in mathematics achievement for a teacher's class was a significant predictor of average status at the 2-year follow-up for all subscales of the Fidelity measure. This suggests that teachers who perceived a greater gain in mathematics within their classroom demonstrated higher levels of fidelity across the study's time frame (consistent with Durlak & DuPre, 2008; Fishman et al., 2011; Sanford DeRousie & Bierman, 2012). Similarly, child gain positively and significantly impacted the rate of growth for both the General Curriculum and Computer subscales. The finding that child outcome scores significantly predicted sustained fidelity to the computer component suggests that teachers who perceived that their children gained more competence in mathematics (recall that we did not inform teachers about the measured gains of individual classes, including their own) increased their effort to implement all of the components of the intervention with greater fidelity. We posit that perceptions of their children's learning may motivate teachers to productively face challenges.

Other factors varied in their effects. Teacher reports of instructional practices had a significant impact on the rate of growth only for the Computer subscale. That is, teachers highly endorsing practices that were promoted in the professional development sessions demonstrated a higher rate of growth in fidelity to the computer component of the intervention. Belief in and implementation of these practices may have supported teachers' commitment to an instructional setting that was particularly different from their previous practices and challenging to implement (instructionally but also technologically). Finally, we examined differences due to amount of teaching experience. Findings indicated that most teachers were willing to implement the innovation, and thus our data do not support either perspective suggested by previous research—that is, that new teachers are either more or less open to such change than more experienced teachers. However, years of experience teaching was related to implementation, with the pattern suggesting that comfort and skill managing the pedagogical contexts may explain the differences. In this instance, the rate of growth for the Small Group subscale was higher for teachers with more experience. Teachers with more experience may have been able to more effectively administer the dual requirements of intense teaching of four children in a small-group setting and management of the rest of the classroom.

Furthermore, growth mixture modeling suggested that these effects differed for different groups of teachers. All teachers who perceived greater mathematics growth in their children had higher levels of fidelity on the General Curriculum subscale, but those in the group who began with higher levels of fidelity on that subscale showed greater growth of fidelity on that subscale. For the Small Group and Computer subscales, the group of teachers that started lower on each of those subscales gained more. Again, this may have been that improvements on the General Curriculum subscale were more likely for those teachers who were at higher levels of competence in all aspects initially and could synthesize disparate elements into a positive classroom culture, whereas for the separate settings (e.g., small group and computer), those teachers with initial lower scores had more room to grow.

Teachers in the high group who reporting greater use of professional development–promoted instructional practices made greater growth than those who reported less frequent use, but the opposite was true for teachers in the low group. Teachers who start out with greater fidelity may become attached to a program earlier and, especially with perceptible gains on the part of their children, become increasingly committed to and skilled with that program. However, those with fewer initial competencies may stumble trying to incorporate new practices (akin to the Matthew effect; Bakermans-Kranenburg, van IJzendoorn, & Bradley, 2005). However, results differed for years of experience. Teachers in the low group who had more years of experience showed greater fidelity growth than those in the low group with fewer years of experience for the Small Group subscale, but the opposite occurred for the Computer subscale. It may be that teachers with greater experience had more experience teaching children in small groups but less experience in and out of school with technology. To the extent that teachers' understanding of the principles of the innovation is exemplified by their self-reports of use of professional development–driven practices, these findings are consistent with the literature showing that teachers' understanding of the principles of the innovation is a predictor of sustainability (cf. Bierman et al., 2013; Sanford DeRousie & Bierman, 2012).

Finally, the whole-group setting was probably the most familiar to teachers and was the setting in which the professional development and experience with the intervention had the least effect, as indexed by the lack of significant change across setting. Unlike the other subscales, teachers in the high group decreased slightly in Whole Group fidelity. Conversely, teachers within the low group demonstrated a positive, although not significant, rate of growth. Both groups came to a central point in Whole Group fidelity similar to the statistically significant, though slight, trajectory for the majority of teachers within the medium group. Child gain was a significant predictor only for the medium group. It may be that most teachers were so familiar with whole-group settings that their teaching strategies from the beginning to the end were well established and routines were therefore resistant to change (Bobis, 2006, p. 16). Also possible is that demands on whole-group time were intense, because both districts were attempting to implement (different) literacy programs in prekindergarten during this period. A limitation of the analyses is the low reliability on the small subscales within the fidelity measure. For the General Curriculum subscale, this is most likely due to the small number of items. Future investigations of fidelity may benefit from the use of more comprehensive subscales.

In summary, the learning trajectory–based instruction engaged in by the prekindergarten teachers in this study may have provided a coherent program of teaching and learning that promoted the significant levels of fidelity found in this study. Teachers taught the curriculum with increasing fidelity as time went by, even though research project staff were no longer able to provide support. They seemed to have internalized the program (Timperley et al., 2007) and to have made sense of the curricular activities involved with whole-group, small-group, and other components within an overall structure of learning trajectories that progressed toward a known mathematical goal. By engaging in the initial professional development, and then becoming empowered by their own knowledge of the trajectories and the ways to support learners through the trajectories, as well as by their perceptions that their students were learning more, they became progressively more faithful to the intended program, instead of drifting from it as time elapsed and support disappeared, as has been documented by other studies (Datnow, 2005; Hargreaves, 2002).

One implication, then, is that a coherent model of professional development, curriculum, instruction, and assessment based on learning trajectories may provide the conditions for promoting sustainability in teacher practices (Clements & Sarama, 2014) and may be particularly beneficial in addressing the climate of low expectations in urban schools (Johnson & Fargo, 2010) as teachers increase their understanding of the capacities of all children to learn mathematics.

As teachers come to understand children's probable developmental paths and become adept at anticipating children's strategies and misconceptions, their teaching practices may become more grounded and solidified. As they monitor for children's multiple models and probe for the ways in which children's mathematical thinking fits the structure of the trajectory, their teaching practices may become reinforced as student reactions provide positive feedback for their practices (Guskey, 2002). Teachers who demonstrate sustained fidelity of implementation to a program that has demonstrated improved child achievement will have a positive impact on many more children than teachers who implement with fidelity only during treatment. Thus, another general implication is that helping teachers perceive and document their children's learning (which human subjects constraints disallowed in this research) may be an effective way of maintaining and even increasing fidelity of implementation. These positive perceptions of learning may be especially important in motivating teachers to productively face the challenges inherent in fully implementing all aspects of the curriculum, which bring their own difficulties. For example, educational technology challenges include limited hardware, hardware and software problems and limited troubleshooting competencies, difficulty scheduling computer use for all children, and inconsistency between computer use and customary practice including contextually constrained choices (Clements & Sarama, 2008; Cuban, 2001). Solving problems successfully, such as engaging children productively in technology activities, may engender confidence and risk taking in future work. Simply, success breeds success, and such changes in practice lead to positive changes in beliefs (e.g., Showers, Joyce, & Bennett, 1987).

Other implications address the various pedagogical contexts. First, administrators, coaches, and trainers may find it effective to help teachers focus on the particular contexts (e.g., whole group, small group, computers) first and then build toward the general classroom context and culture before they synthesize them into a coherent, positive classroom culture for early mathematics. Second, teachers with less experience may benefit from training and coaching focused on unfamiliar grouping strategies, such as small-group instruction, whereas teachers with more experience in the classroom but relatively less experience using computers may benefit from assistance in that domain. Of course, expert coaches may consider such possibilities as initial lenses to use, focusing their observations to individualize their work with individual teachers.

The TRIAD model is not simply about a new curriculum or training teachers to use it. Success required complex changes, including a change in instructional structures, pedagogical strategies, and classroom communication and culture (Grubb, 2008). Given the importance of early competence in mathematics (e.g., Duncan et al., 2007; Paris, Morrison, & Miller, 2006), the TRIAD implementation described here may be important for practice and policy as well as research.

ACKNOWLEDGMENTS

Although this research is concerned with a 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 children who participated in this research.

FUNDING

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305K05157 and R305A110188. The opinions expressed are those of the authors and do not represent the views of the U.S. Department of Education.

REFERENCES

Achinstein, B., & Ogawa, R. T. (2006). (In)fidelity: What the resistance of new teachers reveals about professional principles and prescriptive educational policies. *Harvard Educational Review*, *76*(1), 30–63.

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: Ideas in principle* Vol. 1, (pp. 37–54). Lanham, MD: Rowman & Littlefield.

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Bradley, R. H. (2005). Those who have, receive: The Matthew effect in early childhood intervention in the home environment. *Review of Educational Research*, *75*(1), 1–26. doi:10.3102/00346543075001001

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*(9), 6–8. 14.

Bierman, K. L., Sanford DeRousie, R. M., Heinrichs, B., Domitrovich, C. E., Greenberg, M. T., & Gill, S. (2013). Sustaining high-quality teaching and evidence-based curricula: Follow-up assessment of teachers in the REDI Project. *Early Education & Development*, *24*, 1194–1213. doi:10.1080/10409289.2013.755457

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, *9*(1), 30–52.

Bobis, J. M. (2006). *Factors affecting the sustainability and expansion of Count Me In Too*. Sydney, Australia: New South Wales Department of Education and Training.

Boerst, T., Lambdin, D. V., Confrey, J., White, D., Heck, D., Baltzley, P. C., ... Quander, J. R. (2010). Strengthening research by designing for coherence and connections to practice. *Journal for Research in Mathematics Education*, *41*(3), 216–235.

Burns, M. K., Peters, R., & Noell, G. H. (2008). Using performance feedback to enhance implementation fidelity of the problem-solving team process. *Journal of School Psychology*, *46*, 537–550.

Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E. H., & Empson, S. B. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education*, *29*, 3–20.

Century, J., Rudnick, M., & Freeman, C. (2012, April). *Defining and measuring sustainability of reform: Factors that affect our abilities to generate enduring change*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

Clements, D. H. (2007). Curriculum research: Toward a framework for ''research-based curricula.'' *Journal for Research in Mathematics Education*, *38*, 35–70.

Clements, D. H., & Sarama, J. (2004a). Building Blocks for early childhood mathematics. *Early Childhood Research Quarterly*, *19*, 181–189.

Clements, D. H., & Sarama, J. (2004b). Learning trajectories in mathematics education. *Mathematical Thinking and Learning*, *6*, 81–89. doi:10.1207/s15327833mtl0602_1

Clements, D. H., & Sarama, J. (2007a). *Building Blocks—SRA Real Math teacher's edition, Grade PreK*. Columbus, OH: SRA/McGraw-Hill.

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. (2007c). *SRA Real Math Building Blocks: Teacher's resource guide PreK*. Columbus, OH: SRA/McGraw-Hill.

Clements, D. H., & Sarama, J. (2008). Mathematics and technology: Supporting learning for students and teachers. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on science and technology in early childhood education* (pp. 127–147). Charlotte, NC: Information Age.

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

Clements, D. H., & Sarama, J. (2012). Learning and teaching early and elementary mathematics. In J. S. Carlson & J. R. Levine (Eds.), *Instructional strategies for improving student learning: Focus on early mathematics and reading* Vol. 3, (pp. 107–162). Charlotte, NC: Information Age.

Clements, D. H., & Sarama, J. (2013). *Building Blocks, Volumes 1 and 2*. Columbus, OH: McGraw-Hill Education.

Clements, D. H., & Sarama, J. (2014). Learning trajectories: Foundations for effective, research-based education. In A. P. Maloney, J. Confrey, & K. H. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 1–29). New York, NY: Information Age.

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

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal*, *50*, 812–850. doi:10.3102/0002831212469270

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3–12.

Cuban, L. (2001). *Oversold and underused*. Cambridge, MA: Harvard University Press.

Cuban, L., & Usdan, M. (Eds.). (2003). *Powerful reforms with shallow roots: Improving America's urban schools*. New York, NY: Teachers College Press.

Darrow, C. L. (2013). The effectiveness and precision of intervention fidelity measures in preschool intervention research. *Early Education & Development*, *24*, 1137–1160. doi:10.1080/10409289.2013.765786

Datnow, A. (2005). The sustainability of comprehensive school reform models in changing district and state contexts. *Educational Administrative Quarterly*, *41*(1), 121–153.

Dearing, J. W. (2008). Evolution of diffusion and dissemination theory. *Journal of Public Health Management and Practice*, *14*(2), 99–108.

Dearing, J. W., & Meyer, G. (2006). Revisiting diffusion theory. In A. Singhal & J. W. Dearing (Eds.), *Communication of innovations: A journey with Ev Rogers* (pp. 29–60). New Delhi, India: Sage.

Donaldson, S. I., & Lipsey, M. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The handbook of evaluation: Policies, programs, and practices* (pp. 56–75). London, England: Sage.

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

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350.

Fishman, B. J., Penuel, W. R., Hegedus, S., & Roschelle, J. (2011). What happens when the research ends? Factors related to the sustainability of a technology-infused mathematics curriculum. *Journal of Computers in Mathematics and Science Teaching*, *30*(4), 329–353.

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

Gersten, R., Chard, D. J., & Baker, S. K. (2000). Factors enhancing sustained use of research-based instructional practices. *Journal of Learning Disabilities*, *33*, 445–456.

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

Grubb, W. N. (2008). Multiple resource, multiple outcomes: Testing the ''improved'' school finance with NELS88. *American Educational Research Journal*, *45*, 104–144.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, *8*(3/4), 381–391.

Hargreaves, A. (2002). Sustainability of educational change: The role of social geographies. *Journal of Educational Change*, *3*(3–4), 189–214.

Huberman, M. (1995). Professional careers and professional development: Some intersections. In T. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 193–224). New York, NY: Teachers College Press.

Johnson, C. C., & Fargo, J. D. (2010). Urban school reform enabled by transformative professional development: Impact on teacher change and student learning of science. *Urban Education*, *45*(1), 4–29.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*, 302–317.

Klein, A., Starkey, P., & Wakeley, A. (2000). *Child Math Assessment: Preschool Battery (CMA)*. Berkeley: University of California, Berkeley.

Klingner, J. K., Arguelles, M. E., Hughes, M. T., & Vaughn, S. (2001). Examining the school side ''spread'' of research-based practices. *Learning Disability Quarterly*, *24*(4), 221–234.

Lee, O., Penfield, R., & Maerten-Rivera, J. (2009). Effects of fidelity of implementation on science achievement gains among English language learners. *Journal of Research in Science Teaching*, *46*(7), 836–859.

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

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Authors.

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

National Research Council. (2007). *Taking science to school: Learning and teaching science in Grades K-8*. Washington, DC: National Academies Press.

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

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services & Research*, *3*(4), 374–396. doi:10.1007/s11414-012-9295-x

O'Connell, P. (2012, April). *Is sustainability of educational reform an article of faith, or can it be deliberately crafted?* Paper presented at the annual meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*, 33–84.

Owston, R. (2007). Contextual factors that sustain innovation pedagogical practice using technology: An international study. *Journal of Educational Change*, *8*, 61–77. doi: 10.1007s/10833-006-9006-6

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.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*, 921–958. doi:10.3102/0002831207308221

Pickles, P., & Croudace, T. (2010). Latent mixture models for multivariate and longitudinal outcomes. *Statistical Methods in Medical Research*, *19*, 271–289.

Raudenbush, S. W. (2007). Designing field trials of educational interventions. In B. Schneider & S.-K. McDonald (Eds.), *Scale-up in education: Issues in practice* Vol. 2, (pp. 23–40). Lanham, MD: Rowman & Littlefield.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine*, *22*(2), 237–260.

Sanford DeRousie, R. M., & Bierman, K. L. (2012). Examining the sustainability of an evidence-based preschool curriculum: The REDI program. *Early Childhood Research Quarterly*, *27*(1), 55–56.

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. (2013). Lessons learned in the implementation of the TRIAD scale-up model: Teaching early mathematics with trajectories and technologies. In T. G. Halle, A. J. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 173–191). Baltimore, MD: Brookes.

Sarama, J., Clements, D. H., Henry, J. J., & Swaminathan, S. (1996). Multidisciplinary research perspectives on an implementation of a computer-based mathematics innovation. In E. Jakubowski, D. Watkins, & H. Biske (Eds.), *Proceedings of the 18th annual meeting of the North America Chapter of the International Group for the Psychology of Mathematics Education* Vol. 2, (pp. 560–565). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

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.

Scheirer, M. A. (2005). Is sustainability possible? A review and commentary on empirical studies of program sustainability. *American Journal of Evaluation*, *26*(3), 320–347.

Scheirer, M. A., & Dearing, J. W. (2011). An agenda for research on the sustainability of public health programs. *American Journal of Public Health*, *101*(11), 2059–2067.

Shapley, K. S., Sheehan, D., Maloney, C., & Caranikas-Walker, F. (2010). Evaluating the implementation fidelity of technology immersion and its relationship with student achievement. *Journal of Technology, Learning, and Assessment*, *9*(4), 1–69.

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.

Simon, T. J. (1997). Reconceptualizing the origins of numerical knowledge: A ''nonnumerical'' account. *Cognitive Development*, *12*, 349–372.

Sindelar, P. T., Shearer, D. K., Yendol-Hoppey, D., & Liebert, T. W. (2006). The sustainability of inclusive school reform. *Exceptional Children*, *72*(3), 213–331.

Stein, M. K., Brover, B. W., & Hennigsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classroom. *American Educational Research Journal*, *33*, 455–488.

Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning trajectory based instruction: Toward a theory of teaching. *Educational Researcher*, *41*, 147–156. doi:10.3102/0013189X12442801

Timperley, H. (2011). *Realizing the power of professional learning*. London, England: Open University Press.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional development: Best evidence synthesis iteration*. Wellington, New Zealand: Ministry of Education.

van Driel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. *Journal of Research in Science Teaching*, *38*(2), 137–158.

Varol, F. (2009). *The effects of professional development on preschool teachers' mathematics instruction* (Doctoral dissertation). Vanderbilt University, Nashville, Tennessee.

Vartuli, D., & Rohs, J. (2009). Assurance of outcome evaluation: Curriculum fidelity. *Journal of Research in Childhood Education*, *23*, 502–512.

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

Wu, H.-H. (2011). *Understanding numbers in elementary school mathematics*. Providence, RI: American Mathematical Society.