American Educational Research Journal
June 2008, Vol. 45, No. 2, pp. 443-494
DOI: 10.3102/0002831207312908

Experimental Evaluation of the Effects of a Research-Based Preschool Mathematics Curriculum

Douglas H. Clements
Julie Sarama
University at Buffalo, State University of New York

*A randomized-trials design was used to evaluate the effectiveness of a preschool mathematics program based on a comprehensive model of research-based curricula development. Thirty-six preschool classrooms were assigned to experimental (*Building Blocks*), comparison (a different preschool mathematics curriculum), or control conditions. Children were individually pre- and posttested, participating in 26 weeks of instruction in between. Observational measures indicated that the curricula were implemented with fidelity, and the experimental condition had significant positive effects on classrooms' mathematics environment and teaching. The experimental group score increased significantly more than the comparison group score (effect size = 0.47) and the control group score (effect size = 1.07). Early interventions can increase the quality of the mathematics environment and help preschoolers develop a foundation of mathematics knowledge.*

Keywords: *computers and learning, early childhood, mathematics education, professional development, instructional technologies, curriculum*

Douglas H. Clements is a professor of early childhood, mathematics, and computer education at the University at Buffalo, State University of New York, 505 Baldy Hall (North Campus), Buffalo, NY 14260; e-mail: *clements@buffalo.edu*. His research interests include the learning and teaching of geometry, computer applications in mathematics education, the early development of mathematical ideas, the effects of social interactions on learning, and curriculum research, including the scaling up of effective interventions.

Julie Sarama is an associate professor of mathematics education at the University at Buffalo, State University of New York, 505 Baldy Hall (North Campus), Buffalo, NY 14260; e-mail: *jsarama@buffalo.edu*. 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 her own software environments in mathematics classrooms.

Researchers and government agencies have emphasized the importance of "research-based" instructional materials (e.g., Feuer, Towne, & Shavelson, 2002; Kilpatrick, Swafford, & Findell, 2001; Reeves, 2002), but rigorous evaluations of mathematics curricula are uncommon (National Research Council [NRC], 2004). Rarer are evaluations of preschool mathematics curricula, especially those including children from schools serving low-socioeconomic (SES) children (Clements & Sarama, 2007b). In this study, we used a randomized-trials design to evaluate the effectiveness of a preschool mathematics program based on a comprehensive model of research-based curriculum development. Research issues included the fidelity of implementation, the effects on the quality of the classrooms' mathematics environment and teaching and on preschoolers' mathematics achievement, and the mediational role of the measure of the educational environment on gains in mathematics achievement.

Background and Theoretical Framework

This study was motivated by three related concerns: (a) the need for rigorous evaluations of curricula; (b) the need for preschool mathematics curricula, and evaluations of these curricula, that include children from low-SES backgrounds; and (c) the desire to evaluate instructional materials based on a theoretical model of curriculum development. Regarding the first concern, both the ambiguities of the phrase *research-based instructional materials* and ubiquitous claims that curricula are based on research vitiate attempts to create a research foundation for the creation and evaluation of curricula (Clements, 2007). Once produced, curricula are rarely evaluated scientifically (NRC, 2004; less than 2% of studies address curricula; Senk & Thompson, 2003). Few evaluations of any curricula use randomized field trials (Clements, 2002; NRC, 2004).

Regarding the second concern, although mathematics in preschool has a long history, especially as realized in Froebel's original kindergarten (Balfanz, 1999; Brosterman, 1997), mathematics curricula for preschoolers have not been common, possibly due to the influential position of Piaget that early instruction on number skills would be useless (Piaget & Szeminska, 1952). Traditional preschool curricula often emphasize "prenumber" activities such as classification and seriation, which Piagetian theory identified as cognitive foundations for later number learning (Wright, Stanger, Stafford, & Martland, 2006). However, this approach is less effective than one based on recent research on children's early developing number knowledge (Clements, 1984). The curricula in more recent evaluations, many of which are unpublished materials created by researchers, have focused on mathematics, but most address only a single topic, such as number (Arnold, Fisher, Doctoroff, & Dobbs, 2002; Clements, 1984; Griffin & Case, 1997; Wright et al., 2006) or geometry (Razel & Eylon, 1991). Nevertheless, evaluations suggest that these materials can increase preschoolers' mathematics experiences, strengthening the development of their knowledge of number or geometry (Clements, 1984;

Griffin & Case, 1997; Razel & Eylon, 1991). Few studies have examined the effects of more comprehensive preschool mathematics curricula, and those that exist did not use random assignment or included small numbers of classrooms (e.g., Clements & Sarama, 2004a, 2007c; Klein & Starkey, 2004; Starkey, Klein, & Wakeley, 2004; a combination of components of these curricula was assessed in Institute of Education Sciences' [IES] Preschool Curriculum Evaluation Research program, or PCER, but these results are embargoed). Thus, there is a need for rigorous evaluations of curricula that address comprehensive goals for mathematics learning (National Council of Teachers of Mathematics [NCTM], 2006), especially because learning of different domains, such as number, geometry, measurement, and patterning, may be mutually reinforcing (Clements & Sarama, 2007b).

Further, few rigorous evaluations include children from low-income households, who are at risk for later failure in mathematics (Bowman, Donovan, & Burns, 2001; Denton & West, 2002; Mullis et al., 2000; Natriello, McDill, & Pallas, 1990; Secada, 1992; Starkey & Klein, 1992). These children receive less support for mathematics learning in the home and school environments than children from middle- and high-income households (Blevins-Knabe & Musun-Miller, 1996; Bryant, Burchinal, Lau, & Sparling, 1994; Farran, Silveri, & Culp, 1991; Holloway, Rambaud, Fuller, & Eggers-Pierola, 1995; Saxe, Guberman, & Gearhart, 1987; Starkey et al., 1999). Some of the evaluations of preschool materials have involved low-income children and have suggested that planned and purposeful activities can ameliorate children's lack of experiences and increase their mathematics achievement (e.g., Griffin & Case, 1997; Klein & Starkey, 2004). High-quality evaluations of comprehensive curricula involving such children are needed to guide educational planning and policy.

Most important is the third concern, the need for evaluations of curricula developed on theoretical models. *Building Blocks—Foundations for Mathematical Thinking, Pre-Kindergarten to Grade 2: Research-Based Materials Development* was funded by the National Science Foundation to create and evaluate mathematics curricula for young children based on a theoretically sound research and development framework. The project's curriculum research framework (CRF) includes 10 phases embedded within three categories (Clements, 2002, 2007; for more detail, see Clements & Sarama, 2004a, 2007c; Sarama, 2004). The present research is the first of several Phase 10 evaluations; as background, we briefly describe the first 9 phases, providing citations and examples from the development of *Building Blocks*. The first category, A Priori Foundations, includes three phases that are variants of the research-to-practice model in which extant research is reviewed and implications for the nascent curriculum development effort drawn. (1) In General A Priori Foundation, developers review broad philosophies, theories, and empirical results on learning and teaching (Bowman et al., 2001). We determined that *Building Blocks*' basic approach would be finding the mathematics in, and developing mathematics from, children's activity; for example, "mathematizing" everyday tasks such as setting a table.

(2) In Subject Matter A Priori Foundation, developers review research and consult with experts to identify mathematics that makes a substantive contribution to students' mathematical development, is generative in students' development of future mathematical understanding, and is interesting to students (Clements & Sarama, 2004a). (3) In Pedagogical A Priori Foundation, developers review empirical findings regarding what makes activities educationally effective—motivating and efficacious—to create general guidelines for the generation of activities (Sarama, 2004).

In the second category, Learning Model, developers structure activities in accordance with empirically based models of children's thinking in the targeted subject-matter domain. This phase, (4) Structure According to a Specific Learning Model, is critical to this study; therefore, we will elaborate. We created research-based *learning trajectories* (Simon, 1995), which we define 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" (Clements & Sarama, 2004c, p. 83). For example, children's developmental progression for shape composition (Clements, Wilson, & Sarama, 2004; Sarama, Clements, & Vukelic, 1996) advances through levels of trial and error, partial use of geometric attributes, and mental strategies to synthesize shapes into composite shapes. The sequence of instructional tasks requires children to solve shape puzzles off and on the computer, the structures of which correspond to the levels of this developmental progression (Clements & Sarama, 2007c; Sarama et al., 1996). There is evidence that learning trajectories facilitate all children's learning (Clements & Sarama, 2004b, 2007b; Simon, 1995) but are especially important for children whose development may be attenuated or delayed, because critical levels of thinking may have been missed. For example, although most middle- and high-SES children build strong counting concepts and skills by 3.5 years of age, many low-SES children have not had opportunities to develop one or more critical concepts, such as correspondence, cardinality, or comparison in the counting act (Clements & Sarama, 2007b; Griffin, 2004). Putting learning trajectories at the core ensures that assessment (small-group record sheets, software, etc.) and activities based on it reveal and address any such lacunae. The CRF's model we used for *Building Blocks* implies that topic-specific learning trajectories should be interwoven rather than taught in separate curricular units, for five reasons. First, children's learning of mathematics is continuous and incremental (Clements & Sarama, 2007b; Siegler, 1996), and consistent exposure to skill-building activities is important to children's learning (Fuson, 1988). Second, the progressions for each learning trajectory cover years of development, and compressing the learning into a unit of instruction would be inappropriate. Third, the preschool years is a substantial period of cognitive development, with wide individual differences (Bowman et al., 2001); therefore, distributing opportunities to learn topics across the year will be more effective for more children. Fourth, at all ages, distributed practice yields better recall and retention (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Rohrer & Taylor, 2006). Fifth, interweaving may facilitate mutual reinforcement between learning trajectories (Clements & Sarama, 2007b). As a simple example, early learning of subitizing (rapid recognition of the numerosity of small sets) supports the development of a critical (and oft-neglected) level of thinking in the counting trajectory in which children recognize the last counting word indicates how many in the counted set. That is, if children count a group of objects, "One, two, three," and immediately recognize a group as containing three objects via subitizing, their understanding of the cardinality of the last counting word is facilitated. Conversely, the establishment of that level of thinking in counting supports the development of higher levels of subitizing.

In the third category, Evaluation, developers collect empirical evidence to evaluate appeal, usability, and effectiveness of some version of the curriculum. We conducted studies at each of the next four phases: (5) Market Research; (6) Formative Research: Small Group (pilot tests with small groups on components); (7) Formative Research: Single Classroom; and (8) Formative Research: Multiple Classrooms (e.g., Clements & Sarama, 2004a; Sarama, 2004), revising the curriculum multiple times, including two distinct published versions (Clements & Sarama, 2003, 2007a). Another fundamental way *Building Blocks* was developed to help all learners was to include teachers and children from schools serving low-income families throughout these formative phrases. This helped ensure the problem contexts and language used were appropriate for these populations and that all formative evaluation included empirical data on the effectiveness of activities for supporting these children's learning. In the last two phases, (9) Summative Research: Small Scale and (10) Summative Research: Large Scale, developers evaluate what can actually be achieved with typical teachers under realistic circumstances. An initial Phase 9 summary research project (Clements & Sarama, 2007c) yielded effect sizes between 1 and 2 (Cohen's *d*). Phase 10 also uses randomized trials, which provide the most efficient and least biased designs to assess causal relationships (Cook, 2002), now in a greater number of classrooms, with more diversity, and less ideal conditions.

The present study is the first of several Phase 10 evaluations (Clements, 2007) evaluating the effects of a complete preschool mathematics curriculum on the mathematical knowledge of 4-year-old children, including those attending schools that serve children from low-SES families. Research questions included the following: Can *Building Blocks* be implemented with high fidelity, and does the measure of fidelity predict achievement gains? Does *Building Blocks* have substantial positive effects on the quality of the mathematics environment and teaching? What are the effects of the *Building Blocks* curriculum, as implemented under diverse conditions, on the mathematics achievement of preschoolers? A final, secondary, question was, If these effects are significant, does the increase in the quality of the mathematics environment and teaching mediate the effects on mathematics achievement? The complexity of numerous contexts, compared to the "superrealization" (Cronbach et al., 1980) of the Phase 9 study, along with the small (e.g., .25) to moderate (.5; Cohen, 1977) effect sizes documented for other curricular interventions (NRC, 2004; Riordan & Noyce, 2001) suggested that a reasonable prediction would be moderate to large effect sizes.

Method

Participants

Table 1 presents the population of diverse classrooms serving preschoolers in New York State. The first group, serving children from low-income households, includes Head Start and state-funded programs. From an initial pool of more than 100 volunteers, 24 teachers were randomly selected. The *Building Blocks* curriculum was designed to meet the needs of all children. Therefore, we included the second group, who served mixed- (low- and middle-) income children. From 20 volunteers, 12 teachers were randomly selected. In each classroom, we randomly selected 8 children from the pool of all kindergarten-intending (in the entry range for kindergarten in 2004–2005) preschoolers who returned institutional review board permission forms (a few Head Start classrooms had only 8 kindergarten-intending children who returned forms; in those cases, all those qualifying were tested). One comparison teacher became increasingly ill, and 4 children (3 from control, 1 from *Building Blocks*) moved out of the area during the first 3 months of the study, leaving 35 teachers and 276 children who participated throughout the study (technical problems and children's illnesses during testing resulted in a total of 253 children with complete data on both pretest and posttest). To determine if there was any substantive effect of the teacher leaving the study, we calculated an effect size as the change in mean pretest achievement mean for that condition as a result of the attrition, divided by the pooled standard deviation. The small effect size of .03 indicates that any effect on the findings was negligible.

Materials: Curricula

Researchers should select curricula to which an experimental curriculum is compared on a principled basis (Clements, 2007) and assess the fidelity with which each curriculum is implemented (NRC, 2004). The use of conventional curricula as control conditions is important, but equally useful is the inclusion of a more rigorous comparison curriculum (NRC, 2004). We implemented a research-based comparison curriculum specifically designed for low-income children and validated in previous research with children from low-income households (IES's PCER program).

Table 2 summarizes the curricula's characteristics. For example, the first intervention curriculum, *Building Blocks* (experimental), was developed and evaluated using Phases 1 to 9 of the CRF (with the present study representing Phase 10), as previously described. It typically conducted small-group mathematics sessions once per week for 10 to 15 minutes per session per

Table 1

Demographics of Participating Schools

Program			Teachers						Children				
					Ethnicity (%)					Ethnicity (%)			
Name	Number Pre-Ks	Urbanicity	Experience (Years)	New York State Certification (%)	AA	A/P; NA	H	W	SES (% Free Lunch; % Reduced Lunch)	AA	A/P; NA	H	W
Head Start	9	Urban	8	28	26	1	9	64	97; 2	47	2; 8	13	30
State funded	15	Urban	16	90	19 for AA, A/P, NA, and H			81	63; 11	58	3	11	28
Mixed income	12	Suburban	14	91	5 for AA, A/P, NA, and H			95	9; 10	30 for AA, A/P, NA, and H			70

Note. AA = African American; A/P = Asian/Pacific Islander; NA = Native American; H = Hispanic; W = White non-Hispanic. In some cases, records allowed no further categorical breakdowns.

group of approximately 4 to 6 children and whole-group activities for 5 to 15 minutes about four times per week. Children spent about 10 minutes in computer activities twice per week. In addition, letters describing the mathematics children were learning and family activities that support that learning were sent home each week. *Building Blocks* emphasizes use of learning trajectories.

The second intervention curriculum (comparison) had three components. The main components were included in a mathematics-intensive curriculum, the Preschool Mathematics Curriculum (PMC; Klein, Starkey, & Ramirez, 2002), comprising seven units explicitly linked to the NCTM (2000) standards. The curriculum focuses on small-group activities that were implemented so that each child participated at least twice per week for 15 to 20 minutes per day. These were often introduced during whole-group time; in addition, teachers conducted related mathematics activities during that time, for a total of about 10 minutes per day. The second component of the PMC was parent letters, including family activities. The third component was the DLM Early Childhood Express software, with which children spent 5 to 10 minutes twice per week.

The control teachers continued using their school's mathematics activities, which, typical for preschools, showed a mixture of influences. Five low-income controls used a citywide set of activities and common manipulatives. The other two low-income control classrooms from Head Start used the Creative Curriculum (Teaching Strategies, 2001), including the text and manipulative kit. The mixed-income classrooms used homegrown materials based on state standards, with three employing Montessori mathematics materials. Visits to control classrooms indicated that each was following the curricula as written.

As shown in Table 2, all taught a broad range of mathematical topics using several pedagogical components, with the control conditions being more varied and placing more emphasis on topics such as probability and graphing. All included specifications for individualization. The two intervention curricula shared several features but differed on others. Both were supplemental, mathematics-only curricula whose efficacy was supported by previous research. Weekly dosage was similar. Most differences between the two stemmed from the ways the curricula were based on research. The *Building Blocks* curriculum was, as described previously, based on a comprehensive framework, requiring evidence of success at each formative evaluation phase of the CRF. As opposed to the comparison curriculum's organization into topics, the *Building Blocks* curriculum is structured around interwoven learning trajectories, consistently returning to topics at next higher level of the developmental progression. As opposed to the comparison's small-group activities that were to be followed closely, teachers were to interpret and adapt all activities in the *Building Blocks* curriculum according to their knowledge of the developmental progressions underlying the learning trajectories and their formative assessment of children's knowledge. In the same vein, *Building Blocks* asks teachers to emphasize interaction around children's solution strategies, frequently asking questions such as "How did you know?" and "Why?" because children's responses to such

Table 2
Comparison of Curriculam

Characteristics[a]	Control	Comparison	*Building Blocks*
Phases of the CRF employed	2 for citywide 2, 1, 2, 3, 9, 10 (not rigorous) for Creative Curriculum 1, 2, 3, 9 for Montessori 2 for homegrown	1, 2, 3, 9 (not RCT) for PMC	1 to 9, inclusive
Pedagogical components of curricula (math portion for comprehensive curricula)	Whole group 10 min 4 times/ week (city) Small group 10 min/week (Creative Curriculum only) Individual lessons 5 times/week (Montessori only) Computer varies Family varies	Small group 20 min 2/week Whole group 5 min/week Computer 10 min 2/week Family 1/week	Small group 15 min/week Whole group 10-15 min 4/week Computer 10 min 2/week Family 1/week
Emphasis	Teach required skills through direct instruction and through play (city, homegrown) Individual materials and tasks (Montessori only) Learn math through play, with provided materials and teacher scaffolding (Creative Curriculum)s	Increase informal knowledge using manipulatives in sequenced, topical units, with scripted activities that include scaffolding for lower- and higher-performing children	Curriculum and teaching strategies share a core of interwoven, research-based learning trajectories with activities that were formatively evaluated through the CRF phases

(continued)

Table 2
(continued)

Percentage of Weeks on Topics[b]	Control	Comparison	*Building Blocks*
Number			
Counting (verbal, object, strategies)	20	18	21
Number recognition,matching, subitizing	7	7	7
Comparing number, sequencing	7	7	4
Arithmetic	13	18	14
Geometry			
Shape identification	7	7	11
Comparing shape		7	4
Representing shape	7	4	4
Composing shape	7	7	11
Measurement	13	11	14
Patterning	7	14	7
Other (e.g., probability, graphing, process)	13		

Note. CRF = curriculum research framework; RTC = randomized controlled trial; PMC = preschool mathematics curriculum.

[a]For comparison, and especially control, characteristics were inferred from the existing literature, giving benefit of the doubt for CRF Phases 1, 2, and 3. For control, the most typical characteristics of the various curriculum in use were listed.

[b]Main topic of each week; for many weeks, especially in *Building Blocks*, other topics were integrated. Control classrooms used several curricula; data were calculated from modes.

questions are often requisite to identifying the mathematical strategies used by the child and therefore the developmental level of the learning trajectory.

Measures

Classroom teaching practices and environment. Two observational instruments were designed to be substantial improvements over previous instruments in attempting to address "deep change" that "goes beyond surface structures or procedures (such as changes in materials, classroom organization, or the addition of specific activities) to alter teachers' beliefs, norms of social interaction, and pedagogical principles as enacted in the curriculum" (Coburn, 2003, p. 4). The instruments, Fidelity of Implementation (Fidelity) and Classroom Observation of Early Mathematics–Environment and Teaching (COEMET), were created based on a body of research on the characteristics and teaching strategies of effective teachers of early childhood mathematics (Clarke & Clarke, 2004; Clements & Conference Working Group, 2004; Fraivillig, Murphy, & Fuson, 1999; Galván Carlan, 2000; Galván Carlan & Copley, 2000; Horizon Research Inc., 2001; NAEYC, 1991; Teaching Strategies, 2001). Each item is connected to one or more of these studies; thus, there is intended overlap between the instruments, with each specialized for its purpose. An example of a Likert item shared by both instruments in the section Mathematical Focus, with response possibilities from *strongly disagree* to *strongly agree*, 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)." Also shared by both instruments in the section for an interactive mathematics activity titled Organization, Teaching Approaches, Interactions are items with the subheadings Expectations, Eliciting Children's Solution Methods, Supporting Children's Conceptual Understanding, and so forth. Thus, although the fidelity instrument includes additional items measuring compliance, both the Fidelity and COEMET instruments were designed to more deeply document how mathematics is taught and what happens in each classroom (Hall & Hord, 2001).

The Fidelity instrument evaluates the degree to which teachers taught the intervention curricula, thus it addresses adherence and integrity to a specific program but is sufficiently general to apply to either of the two specific intervention curricula. There are 61 items, all but 6 of which are 4-point Likert scales from *strongly disagree* (1) to *strongly agree* (4). As with all measures in this study, we submitted this instrument to the Rasch model, with scores converted to T scores (M = 50, SD = 10). The Rasch T score for Fidelity includes the 55 Likert items and six additional variables: number of adults in the room, number of whole group activities, and duration of activities. An example of an item unique to the Fidelity measure in the Organization, Teaching Approaches, Interactions section is "The teacher conducted the activity as written in the curriculum, or made positive adaptations to it (not changes that violated the spirit of the core mathematical activity)." Further, as shown in Table 3, the Fidelity instrument includes sections for

(text continues on p. 461)

Table 3

Means and Standards Deviations for the Fidelity of Implementation Measure

	Comparison				*Building Blocks*			
	Observation			Mean	Observation			Mean
	1	2	3	3 Obs.	1	2	3	3 Obs.
Mean *T* score[a]	51.0 (9.9)	50.7 (8.8)	48.5 (15.5)	50.1 (10.7)	50.0 (7.6)	52.3 (9.7)	50.8 (9.1)	51.0 (6.6)
General Curriculum								
Schedule	3.6	3.1	1.9	2.9	3.6	3.1	3.4	3.4
Teacher within weeks of schedule[b]	3.6	3.1	1.9	2.9	3.6	3.1	3.4	3.4
Family involvement	3.6	3.6	2.7	3.3	3.8	3.8	4.0	3.9
Activities were sent home[b]	3.6	3.6	2.7	3.3	3.8	3.8	4.0	3.9
Everyday activities	3.4	3.3	3.4	3.4	3.3	3.4	3.1	3.2
Materials were present	3.1	3.3	3.4	3.3	3.4	3.4	3.5	3.4
Teacher uses curriculum every day	3.7	3.3	3.3	3.4	3.1	3.4	2.6	3.0
Extensions	2.7	2.7	2.6	2.7	2.8	3.1	2.6	2.8
Teacher extended activities	2.7	2.7	2.6	2.7	2.8	3.1	2.6	2.8
Whole-Group Activity								
Mathematical focus	5.6	4.7	5.6	5.3	3.9	4.5	4.3	4.2
Teacher displayed understanding of concepts	5.6	4.7	5.6	5.3	3.9	4.5	4.3	4.2
Organization, teaching, approaches, interactions	5.1	4.8	4.9	4.9	4.0	4.2	3.9	4.0
Materials set up correctly	5.9	5.7	5.8	5.8	4.9	4.9	5.1	5.0
Teacher began by focusing thinking	4.1	4.6	5.0	4.6	3.4	4.4	3.9	3.9
Pace was appropriate	5.4	4.9	5.0	5.1	4.4	4.4	4.4	4.4
Teacher conducted activity as written	5.3	5.1	5.0	5.1	4.3	4.3	4.2	4.3
Management strategies enhanced quality	5.7	4.4	5.4	5.2	3.9	4.2	3.6	3.9
Discussion	4.1	4.0	3.4	3.8	3.1	2.8	2.4	2.8
Activity involved discussion	4.1	4.0	3.4	3.8	3.1	2.8	2.4	2.8
Small-Group Activity								
Mathematical focus	3.3	3.4	3.8	3.5	3.3	3.0	3.6	3.3
Teacher displayed understanding of concepts	3.3	3.4	3.8	3.5	3.3	3.0	3.6	3.3

(continued)

Table 3
(continued)

	Comparison				*Building Blocks*			
	Observation			Mean	Observation			Mean
	1	2	3	3 Obs.	1	2	3	3 Obs.
Organization, teaching, approaches, interactions	3.2	3.0	3.2	3.1	3.1	3.2	3.3	3.2
Materials were set up correctly[b]	4.0	4.0	4.0	4.0	3.8	4.0	3.8	3.9
Teacher conducted activity as written	3.7	4.0	3.3	3.7	2.9	3.1	3.5	3.2
Pace was appropriate	3.6	3.4	3.7	3.6	3.6	3.3	3.4	3.4
Activity was completed with all children[b]	4.0	2.7	3.0	3.2	3.6	3.8	4.0	3.8
Management strategies high quality	3.6	3.0	3.2	3.3	3.4	3.0	3.1	3.2
Expectations	2.9	2.8	3.4	3.0	3.1	3.0	3.3	3.1
Teacher promoted effort, persistence	3.4	3.1	3.5	3.3	3.3	3.1	3.4	3.3
Teacher encouraged active reflection	2.3	2.4	3.2	2.6	2.9	2.9	3.1	3.0
Eliciting children's solution methods	2.5	2.5	2.7	2.5	2.8	2.9	2.9	2.9
Teacher asked children to share, justify	2.6	2.4	2.5	2.5	2.8	3.0	3.1	3.0
Teacher facilitated children's responding	2.6	3.0	3.0	2.9	3.4	3.3	3.1	3.3
Teacher encouraged children's listening	2.3	2.0	2.5	2.3	2.3	2.5	2.6	2.5
Supporting children's conceptual understanding	2.4	2.6	2.8	2.6	2.5	2.7	2.9	2.7
Supported describer's thinking	2.4	2.7	3.3	2.8	2.6	2.9	3.1	2.9
Supported listener's understanding	1.9	2.1	2.2	2.1	2.0	2.4	2.5	2.3
Gave just enough assistance	2.9	2.9	2.8	2.9	2.9	2.9	3.1	3.0
Extending children's mathematical thinking	2.2	2.3	2.9	2.5	2.1	2.5	2.7	2.4
Elaborated children's mathematical ideas	2.0	2.1	3.0	2.4	2.1	2.6	2.9	2.5
Went beyond initial solutions	1.6	1.9	2.2	1.9	1.7	2.2	2.3	2.1

(continued)

Table 3
(continued)

	Comparison				*Building Blocks*			
	Observation			Mean	Observation			Mean
	1	2	3	3 Obs.	1	2	3	3 Obs.
Encouraged mathematical reflection	2.4	2.6	3.2	2.7	1.8	2.4	2.6	2.3
Cultivated love of challenge	2.6	2.6	3.2	2.8	2.6	2.8	3.0	2.8
Assessment and instructional adjustment	3.1	3.3	3.1	3.2	2.9	3.0	3.0	3.0
Listened to children, taking notes	3.7	3.6	3.7	3.7	3.4	3.4	3.4	3.4
Adapted tasks to ability and development	2.7	3.0	3.0	2.9	2.8	3.0	2.9	2.9
Used scaffolding activities	3.4	3.4	3.2	3.3	2.6	2.6	2.6	2.6
Used upward and downward extensions[b]	2.7	3.1	2.5	2.8	2.7	2.9	3.1	2.9
Center Activity								
Organization, teaching, approaches, interactions	3.5	3.2	4.3	3.6	3.3	4.0	3.3	3.5
Tasks engaged children	3.6	3.2	4.4	3.7	3.3	4.1	3.3	3.6
Task selected by child[b]	3.4	2.2	3.2	2.9	3.4	3.8	3.8	3.7
Materials set up correctly[b]	4.0	3.4	5.0	4.1	3.8	4.2	4.0	4.0
Teacher introduced activity as written	4.0	3.6	4.8	4.1	3.1	4.1	2.9	3.4
Teacher guided as needed	3.0	3.4	4.0	3.5	3.3	4.1	2.9	3.4
Management strategies enhanced quality	2.8	3.2	4.4	3.5	3.0	3.4	2.7	3.0
Computer Activity								
Organization, teaching, approaches, interactions	2.9	2.8	2.7	2.8	3.0	3.1	2.7	2.9
Materials were set up correctly[b]	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Child was "signed in" with correct name[b]	4.0	4.0	4.0	4.0	3.8	4.0	4.0	3.9
Teacher focused mathematical thinking	1.6	2.6	2.4	2.2	2.2	2.2	2.1	2.2
Teacher monitored, was available as needed	3.1	3.6	2.7	3.1	3.1	3.1	2.8	3.0
Management strategies enhanced quality	3.3	2.9	2.7	3.0	3.0	2.6	2.5	2.7
All children engaged in activity that week	3.1	3.1	3.1	3.1	3.6	3.8	3.1	3.5

(continued)

Table 3
(continued)

	Comparison				*Building Blocks*			
	Observation			Mean	Observation			Mean
	1	2	3	3 Obs.	1	2	3	3 Obs.
Teacher was actively involved	2.3	1.4	1.9	1.9	1.6	1.2	1.6	1.5
Percentage time teacher actively involved	2.0	1.0	1.0	1.3	2.7	4.0	1.7	2.8
Mathematical focus	2.4	4.0	4.0	3.5	3.3	4.0	3.8	3.7
Teaching strategies appropriate	2.7	4.0	4.0	3.6	3.0	4.0	3.7	3.6
Expectations	2.5	4.0	4.0	3.5	3.7	4.0	3.9	3.9
High, realistic expectations	3.0	4.0	4.0	3.7	3.7	4.0	3.7	3.8
Teacher promoted effort	2.0	4.0	4.0	3.3	3.7	4.0	4.0	3.9
Supporting children's conceptual understanding	2.0	4.0	4.0	3.3	3.3	4.0	3.7	3.7
Teacher gave just enough assistance	2.0	4.0	4.0	3.3	3.3	4.0	3.7	3.7
Assessment and instructional adjustment	1.9	3.5	3.5	3.0	3.9	4.0	3.9	3.9
Teacher monitored activity, taking notes	1.7	3.0	3.0	2.6	3.7	4.0	3.7	3.8
Teacher can access records[b]	2.0	4.0	4.0	3.3	4.0	4.0	4.0	4.0
Descriptive Items								
Total adults in classroom	2.0	2.1	2.1	2.1	2.1	2.1	2.0	2.1
Number of whole-group activities	1.6	1.4	1.1	1.4	1.2	1.3	1.3	1.3
Total whole-group duration (up to 3 activities)[d, e]	3.4	3.3	2.3	3.0	2.6	2.6	2.4	2.5
Small-group duration[d]	21.7	20.0	11.3	17.7	15.7	14.3	11.9	14.0
Number of center activities	0.7	0.7	1.0	0.8	1.0	1.2	1.0	1.1
Total center duration (up to 3 activities)[d, f]	1.1	1.3	2.0	1.5	2.2	2.9	2.6	2.6
Computer activity duration[d]	38.7	30.3	49.0	39.3	46.4	52.3	53.1	50.6

Note. Comparison $n = 7$; *Building Blocks* $n = 14$. Scale for all items without Note b or c: 1 = *strongly disagree*, 2 = *disagree*, 3 = *agree*, 4 = *strongly agree*.

[a]Standard deviation in parentheses.

[b]1 = no; 4 = yes.

[c]Scale for percentage items: 1 = 0% to 24%; 2 = 25% to 49%; 3 = 50% to 74%; 4 = 75% to 100%.

[d]In minutes.

[e]1 to 5 minutes = 1; 6 to 9 minutes = 2; 10+ minutes = 3.

[f]1 to 20 minutes = 1; 21 to 45 minutes = 2; 46+ minutes = 3.

Table 4
Means and Standards Deviations for the Classroom Observations (COEMET)

		Control				Comparison				*Building Blocks*			
		Observation			*M*	Observation			*M*	Observation			*M*
Scored Items	Diff.	1	2	3	3 Obs.	1	2	3	3 Obs.	1	2	3	3 Obs
T-score mean		49.7	45.4	43.4	46.2	52.2	47.1	47.7	49.1	57.0	55.7	53.7	55.5
SD		11.6	7.5	8.3	7.3	9.5	6.4	9.2	5.1	13.2	6.0	11.0	7.6
Classroom elements													
Number of computers running math activities	50.8	1.1	0.9	1.6	1.2	1.6	2.0	1.4	1.7	1.9	2.4	2.4	2.2
Mean number of math activities used	53.2	3.6	3.2	3.4	3.4	3.3	3.1	2.6	3.0	5.5	5.3	4.8	5.2
Mean duration of math activities[a]	63.2	22.6	12.3	14.4	16.4	24.7	16.1	16.0	18.9	21.9	19.0	13.9	18.3
Percentage teachers stayed in classroom	33.4	78.6	100	85.7	88.1	100	100	85.7	95.2	100	92.9	92.9	95.3
Classroom culture		2.8	2.9	2.7	2.8	3.5	3.5	3.4	3.5	3.2	3.3	3.0	3.8
Environment and interaction		2.4	2.9	2.3	2.6	3.1	3.4	3.2	3.2	3.1	3.2	2.8	3.0
Interacted with children	25.8	3.6	3.7	3.6	3.7	4.0	4.0	4.0	4.0	3.9	3.9	3.9	3.9
Used teachable moments	50.4	2.9	2.9	2.4	2.7	3.56	3.1	3.4	3.4	3.2	2.9	2.9	2.9
Percentage time use computers	47.5	1.4	3.1	1.4	1.9	3.3	3.7	2.4	3.1	2.7	3.6	2.2	2.8
Math work displayed	52.8	1.7	2.1	1.9	1.9	1.7	2.7	2.9	2.4	2.9	2.5	2.2	2.5
Personal attributes of the teacher		3.2	2.9	3.0	3.0	3.9	3.5	3.6	3.7	3.2	3.4	3.3	3.3
Knowledgeable about math	43.2	3.3	2.9	3.1	3.1	4.0	3.4	3.6	3.7	3.3	3.5	3.4	3.4
Believed math learning enjoyable	43.1	3.2	3.0	3.0	3.1	4.0	3.6	3.6	3.7	3.2	3.4	3.4	3.3

(continued)

Table 4
(continued)

		Control				Comparison				*Building Blocks*			
		Observation			*M*	Observation			*M*	Observation			*M*
Scored Items	Diff.	1	2	3	3 Obs.	1	2	3	3 Obs.	1	2	3	3 Obs.
Enthusiasm for math ideas	45.8	3.1	2.7	2.9	2.9	3.7	3.6	3.6	3.6	3.1	3.4	3.1	3.2
Specific math activities		12.2	8.5	7.9	9.5	10.6	8.6	8.5	9.2	17.2	15.0	14.3	15.5
Mathematical focus		14.3	9.6	10.6	11.5	12.4	10.0	9.9	10.7	21.5	18.3	17.3	19.0
Understanding of topic	55.8	14.1	9.4	10.6	11.3	12.3	10.6	9.7	10.9	20.7	17.7	17.0	18.5
Developmentally appropriate	55.5	14.5	9.9	10.6	11.6	12.4	9.4	10.0	10.6	22.2	18.8	17.5	19.5
Organization, approaches, interactions	11.8	8.4	7.4	9.2	10.3	8.4	8.6	9.1	17.1	14.5	13.4	15.0	
Engaged children's math thinking	62.0	12.2	8.5	7.1	9.3	10.4	9.4	9.0	9.6	18.4	16.9	15.6	17.0
Pace appropriate	57.6	14.1	10.1	10.6	11.6	12.0	10.1	9.7	10.6	21.8	17.3	16.0	18.4
Management strategies	57.2	13.8	9.7	10.1	11.2	11.1	10.0	8.0	9.7	19.4	16.4	13.4	16.4
Actively involved	54.6	13.8	10.5	8.7	11.0	11.3	10.9	10.3	10.8	20.5	18.4	15.9	18.2
Percentage time involved	57.9	10.7	8.4	6.6	8.6	10.4	10.1	10.3	10.3	17.4	15.9	14.6	16.0
Appropriate strategies	63.2	11.9	7.5	6.8	8.7	10.6	8.1	9.0	9.2	17.0	13.8	13.0	14.6
Expectations	11.8	8.4	6.8	9.0	10.3	7.1	8.6	8.7	16.9	13.8	12.9	14.6	
High, realistic expectations	63.9	11.7	8.7	6.7	9.0	10.6	6.7	8.6	8.6	16.6	13.5	12.2	14.1
Acknowledged effort	63.2	11.8	8.2	6.8	8.9	10.0	7.4	8.7	8.7	17.2	14.1	13.6	15.0
Eliciting children's solution methods		10.0	7.1	5.9	7.7	8.8	6.4	7.8	7.7	13.3	11.5	11.7	12.1
Asked children to share ideas	60.1	9.8	7.1	6.1	7.7	8.7	6.6	7.3	7.5	13.5	11.5	11.1	12.0
Facilitated children's responses	58.3	10.7	7.2	5.8	7.9	9.1	7.3	8.4	8.3	13.9	11.9	11.8	12.5
Encouraged evaluating others	60.8	9.5	7.0	5.7	7.4	8.6	5.3	7.6	7.1	12.5	10.9	12.2	11.9

(continued)

Table 4
(continued)

		Control				Comparison				*Building Blocks*			
		Observation			*M*	Observation			*M*	Observation			*M*
Scored Items	Diff.	1	2	3	3 Obs.	1	2	3	3 Obs.	1	2	3	3 Obs.
Supporting children's conceptual understanding		10.3	6.9	5.4	7.5	9.2	6.3	6.7	7.4	13.6	10.9	11.2	11.9
Supported describer's thinking	59.5	10.9	6.6	5.1	7.5	9.1	6.0	7.3	7.5	13.3	10.6	11.2	11.7
Supported listener's understanding	62.5	9.1	6.0	4.5	6.5	8.6	5.6	4.2	6.1	12.3	10.1	10.1	10.8
Gave just enough assistance	64.9	11.0	8.2	6.6	8.6	10.0	7.3	8.7	8.7	15.2	12.1	12.4	13.2
Extending children's math thinking		9.7	6.9	5.7	7.4	8.5	5.7	6.7	7.0	12.2	9.7	9.4	10.4
Elaborated children's ideas	67.5	9.7	7.2	5.6	7.5	8.3	5.4	6.7	6.8	12.8	10.0	10.3	11.1
Encouraged reflection	62.0	9.6	6.6	5.8	7.3	8.7	6.0	6.7	7.1	11.5	9.4	8.5	9.8
Formative assessment		10.4	7.4	5.8	7.9	9.2	7.4	7.0	7.8	13.2	12.3	12.3	12.6
Listened to children, taking notes	65.4	10.0	7.5	5.9	7.8	9.3	8.0	8.3	8.5	12.6	12.6	12.8	12.7
Adapted activities and subtasks to ability and development	66.3	10.8	7.3	5.7	7.9	9.1	6.7	5.7	7.2	13.8	13.1	13.9	13.6
Nonscored Items													
Number of children		16.1	15.4	16.8	16.1	15.1	15.0	15.1	15.1	14.6	14.6	14.6	14.6
Number volunteers		0.1	0.2	0.3	0.2	0.3	0.0	0.0	0.1	0.0	0.2	0.0	0.1
Number of adults		2.3	2.2	2.0	2.2	2.3	2.1	1.3	1.9	2.1	2.3	2.1	2.1

Note. Diff. = Rasch item difficulty expressed as *T* score. For classroom culture and specific math activities, scale for percentage items is 1 = 0% to 24%; 2 = 25% to 49%; 3 = 50% to 74%; 4 = 75% to 100%. Scale for all other items: 1 = *strongly disagree*, 2 = *disagree*, 3 = *agree*, 4 = *strongly agree*, summed across all specific math activities. Control and *Building Blocks*, n = 14; Comparison, n = 7. COEMET = Classroom Observation of Early Mathematics–Environment and Teaching.

[a]In minutes.

each component of the implemented curriculum, such as a specific small-group or family activity. Only activities prescribed in the curriculum implemented are evaluated, and ratings are conducted in reference to the printed curriculum (details of which assessors must be well informed). To see an activity from each component of each curriculum, visits are usually approximately an hour's duration. Interrater reliability, computed via simultaneous classroom visits by pairs of observers (10% of all observations, with pair memberships rotated), averaged 91%. Rasch model reliability is .90.

The COEMET instrument measures the quality of the mathematics environment and activities with an observation of 3 or more hours and is not connected to any curriculum. Thus, it allows for intervention-control condition contrasts, no matter what the source of the enacted curriculum. There are 31 items, all but 4 of which are 4-point Likert scales. An example of one of the three items in a section unique to this measure, 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)." Assessors spend no less than a half day in the classroom, for example, from before the children arrive until the end of the half day (e.g., until lunch). All mathematics activities are observed and evaluated, without reference to any printed curriculum (i.e., assessors are not told what curriculum is present). As shown in Table 4, the COEMET has three main sections, Classroom Elements, Classroom Culture, and Specific Math Activities (SMA). Assessors complete the first two sections once to reflect their entire observation. They complete a SMA form for each observed math activity, defined as one conducted intentionally by the teacher involving several interactions with one or more children or set up or conducted intentionally to develop mathematics knowledge (this would not include, for instance, a single, informal comment). Interrater reliability for the COEMET, computed via simultaneous classroom visits by pairs of observers (10% of all observations, with pair memberships rotated), is 88%; 99% of the disagreements were the same polarity (i.e., if one was *agree*, the other was *strongly agree*). Coefficient alpha (interitem correlations) for the two instruments ranged from .95 to .97 in previous research. Rasch model reliability is .96 for the COEMET.

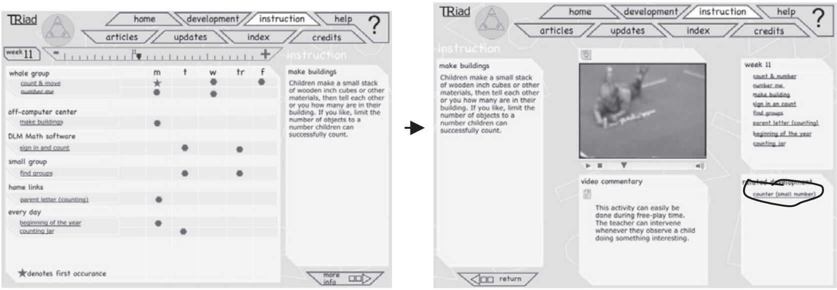
Children's mathematical knowledge. The third instrument measured children's mathematical knowledge and skills. Other instruments were deemed too limited in coverage (e.g., restricted topics, usually only number, and restricted range, such as the Woodcock-Johnson's multiple tasks on numbers 1 to 4). No available instruments avoided these limitations, according to two national panels on preschool assessment (NICHD Forum, Washington, DC, June 2002; CIRCL Forum, Temple University, January 30–31, 2003). Thus, we used the Early Mathematics Assessment (EMA), a measure of preschool children's mathematical knowledge and skills that features two individual interviews of each child of about 10 to 20 minutes, with explicit protocol, coding, and scoring procedures. All sessions are videotaped, and each item is coded for accuracy and, when relevant, for solution strategy used by two trained coders. Any discrepancies are resolved via consultation with the senior researchers. The EMA assesses children's development in a comprehensive set of mathematical topics (see Table 5), rather than mirroring any curriculum's objectives or activities (the EMA was developed before the *Building Blocks* curriculum materials and covers topics such as measurement that are not emphasized in that curriculum). All items are ordered by Rasch item difficulty; children stop the assessment after four consecutive errors. Content validity was assessed via expert panel review; concurrent validity was established with a .86 correlation with a separate research-based instrument (Klein, Starkey, & Wakeley, 2000). The assessment was refined in three pilot tests (Clements & Sarama, 2007c), and a Rasch model analysis was computed, yielding a reliability of .94 for a similar population of children (Sarama & Clements, in press); on the present population, the reliability was .93. For the present study, inferential statistics were performed on Rasch scores computed on correctness scores for the total instrument (and logits transformed to *T* scores, $M = 50$, $SD = 10$, for ease of interpretation). In addition, the sum of raw scores (1 = *correct*, 0 = *incorrect*) was computed for items within each mathematical topic for descriptive purposes.

Procedure

Procedures involved five categories of actions, including initial preparation (e.g., training of assessors and coaches), random assignment of classrooms, teacher training and curriculum implementation, data collection, and analyses.

Initial preparation. We planned, in addition to direct training and practice with the curricula, that intervention teachers would receive in-class coaching. Therefore, local teachers experienced in teaching the curriculum were trained to be coaches during the summer. Also during the summer, assessors were trained on their respective instruments until they achieved 100% accuracy on following the protocol. Graduate students in educational psychology were trained on the EMA, assessing children not in the study, and had to be certified by submitting three consecutive videotapes documenting error-free administration. They remained naïve to children's assigned condition. Members of the *Building Blocks* project staff served as mentors, responsible for administering the Fidelity instrument and monitoring both teachers' implementation and coaches' interventions. Retired teachers, identified by administrators as expert in early childhood mathematics teaching, were trained on the COEMET and were naïve to the classrooms' condition and, further, were unfamiliar with either of the intervention curricula. Thus, those working with the experimental

BBLT **provides scalable access to the learning trajectories via descriptions, videos, and commentaries. Each aspect of the learning trajectories—*developmental progressions* of children's thinking and connected *instruction*—are linked to the other. For example, teachers might choose the instruction (curriculum) view and see the screen on the left, below. Clicking on a specific activity provides a description. Clicking on more info slides the screen over to reveal descriptions, several videos of the activity "in action," notes on the video, and the level of thinking in the learning trajectory that activity is designed to develop, as shown below on the right. (See UBTRIAD.org for a demonstration.)**



Clicking on the related developmental level, or child's level of thinking, ringed above, switches to the development view *of that topic and that level of thinking.* This likewise provides a description, video, and commentary on the developmental level—the video here is of a clinical interview task in which a child displays that *level of thinking.* Teachers can also study a development view, studying clinical interviews of children at each level of thinking, and, if desired, link back to activities.

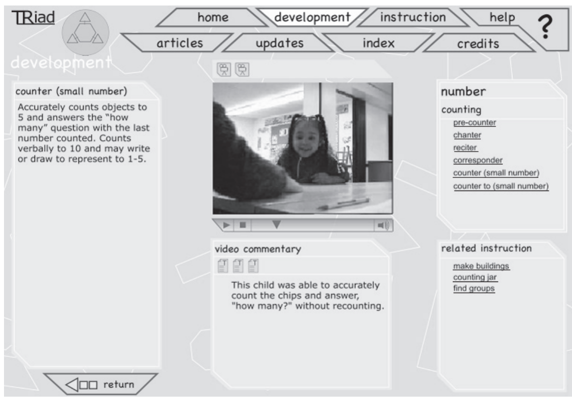


Figure 1. **Description of the *Building Blocks Learning Trajectories* (*BBLT*) Internet-based application.**

teachers on achieving fidelity, and assessing fidelity, were familiar with the curriculum and with teachers' assigned conditions, whereas those assessing both experimental and control conditions were naïve to condition.

Assignment of classrooms. The 24 low-income classrooms were publicly (in presence of four staff members and two school administrators) and randomly (using a table of random numbers, with blind pointing to establish the starting number) assigned to one of three conditions: *Building Blocks*, comparison, or control (one comparison teacher left the area in midfall, leaving 7 classrooms assigned to that condition). The mixed-income classrooms similarly were randomly assigned to *Building Blocks* or control conditions.

Teacher training and curriculum implementation. Teachers in both intervention groups received training, including 4 days and 2-hour refresher classes once every other month. Both groups addressed the following topics but always in the context of the specific curriculum to which they were assigned: supporting mathematical development in the classroom, recognizing and supporting mathematics throughout the day, setting up mathematics learning centers, teaching with computers (including use of the management system and research-based teaching strategies), small-group activities, and supporting mathematical development in the home.

Consistent with the curriculum, only the *Building Blocks* training focused on learning trajectories, such as using learning trajectories for formative assessment. A central tool to support teachers' understanding of learning trajectories, including the goal, the developmental progression of children's thinking, and correlated instructional tasks, was a Web-based application, *Building Blocks* Learning Trajectories. This application provides scalable access to the learning trajectories via descriptions, videos, and commentaries of both the developmental progressions of children's thinking and instruction (see Figure 1). This focus on learning trajectories resulted in the *Building Blocks* group spending less time than the comparison group practicing the curriculum's activities in pairs.

Finally, training for both groups included monthly in-class coaching. Coaching included monitoring, reinforcing, suggesting alternatives, and collaborative problem solving, emphasizing only one or two issues per visit and focusing on implementation of the specific curriculum. Coaching reminds teachers that the project is a priority, that a commitment has been made to it, and that somebody cares about them (Hord, Rutherford, Huling-Austin, & Hall, 1987).

All intervention (comparison and *Building Blocks*) teachers participated in all training activities and implemented their respective curriculum. Control teachers taught the curriculum they had used the year before and agreed to participate in all the data collection (they received the same teacher training as the intervention teachers received the year following the data collection). Participating teachers maintained their daily activities and schedule, including circle (whole-group) time, work at centers, snack, outdoor play, and so forth. The intervention teachers merely inserted the mathematics activities at the appropriate point of the day. For example, in *Building Blocks* classrooms, circle time might include a finger play involving counting, a whole-group counting activity, and an introduction to a new mathematics center or game. Teachers led small-group activities and children worked on the computer activities individually during center time. The comparison classrooms followed similar procedures, but they emphasized small-group activities.

Data collection. Children in all classrooms were assessed at the beginning and end of the school year using the EMA. Teachers began teaching mathematics after the beginning assessments were completed. Mentors collected fidelity data in the intervention classrooms in three time periods:

early fall (after mathematics instruction had begun), winter, and late spring. COEMET observers similarly collected three times during the year.

Analyses. Factorial repeated measures analyses were conducted on the Fidelity (intervention groups) and COEMET (all groups) *T* scores. Standardized mean difference effect sizes, estimates of Cohen's *d*, were computed using the following formula (Lipsey & Wilson, 2001):

$$ES = \frac{M_t - M_c}{\sqrt{\frac{((n_t - 1) \bullet \sigma_t^2) + ((n_c - 1) \bullet \sigma_c^2)}{(n_t - 1) + (n_c - 1)}}}$$

where

$M_t - M_c$ is the difference between the mean gains of the treatment and comparison/control groups,
n_t and n_c are the number of children in the treatment group and comparison/control groups, respectively, and
σ_t and σ_c are the posttest standard deviations of the treatment and comparison/control groups, respectively.

Child mathematics outcomes were coded and then scored by trained teams (not the assessors) naïve to the children's treatment group. Because children were nested within classrooms, child outcome data were analyzed using hierarchical linear models (HLM). This study was a cluster randomized trial, with the classroom the unit of random assignment; HLM accounts for both child- and classroom-level sources of variability in outcomes by specifying a two-level hierarchical model (Raudenbush, 1997). Thus, two-level analyses on the Rasch scores were computed to assess the effectiveness of the curricula and to ascertain the effects of class-level (Level 2) and child-level (Level 1) predictors and interactions of those predictors with treatment group. The Level 1 model was

$$Y_{ij} = \beta_{0j} + r_{ij},$$

where
Y_{ij} is the gain in latent mathematical competence of child i in class j (j = 1 . . . 35 classrooms),
β_{0j} is the mean outcome in class j, and
r_{ij} is the residual (Level 1 random effect).

The Level 2 model was

$$\beta_{0j} = \gamma_{00j} + \gamma_{01}(ClSES)_j + \gamma_{02}(PT)_j + \gamma_{03}(BB)_j + \gamma_{04}(Cmpr)_j + \gamma_{05}(iPTBB)_j + \gamma_{06}(iPTCmpr)_j + \gamma_{07}(iClSESBB)_j + u_{0j},$$

where

γ_{00} is the mean achsievement in the classrooms (intercept),
ClSES is a dummy code for low- or mixed-SES classrooms,
γ_{01} is the main effect for class SES,

PT is a dummy code for program type (Head Start or state funded),
γ_{02} is the main effect for program type,
BB is a treatment-indicator variable for *Building Blocks*,
γ_{03} is the treatment effect for *Building Blocks*,
Cmpr is a treatment-indicator variable for the comparison treatment,
γ_{04} is the treatment effect for the comparison treatment,
iPTBB is the interaction of *PT* and *BB*,
γ_{05} is that interaction effect,
iPTCmpr is the interaction of *PT* and *Cmpr*,
γ_{06} is that interaction effect,
iClSESBB is the interaction of *ClSES* and *BB*,
γ_{07} is that interaction effect, and
u_{0j} is the residual (Level 2 random effect).

All Level 2 predictors were centered around their grand means. All interactions were computed on mean-centered transformations of the variables involved. Effect sizes were computed for significant main effects by dividing the regression coefficient by the pooled posttest standard deviation (for comparison purposes, we also computed *ES* using the previously defined formula for standardized mean difference effect sizes to child-level scores).

Finally, the posttest EMA score was regressed on the COEMET after controlling for EMA pretest score to test whether the observations predicted children's learning. A multiple-regression approach was used to estimate the mediational model (Baron & Kenny, 1986). A series of three regression equations were estimated: (a) We regressed the mediator (COEMET) on the independent variable (treatment group); (b) we regressed the dependent variable (children's gain in mathematics achievement) on the independent variable (treatment); (c) we regressed the dependent variable (gain) on the independent variable (treatment) and the mediator (COEMET), with the mediator entered first. The mediational hypothesis requires that all three equations account for a significant amount of the variance and that when variations in the mediator are controlled, the strength of the previously significant relationship between the independent and dependent variables decreases. Strong evidence for mediation is provided when the relationship between the independent and dependent variables is reduced to zero, but given multiply determined phenomena in social sciences, reducing the relationship constitutes realistic evidence for partial mediation. An alpha level of .05 was used for all statistical tests.

Results

Fidelity

To measure whether the intervention curricula were implemented with fidelity, descriptive statistics were computed. Table 3 shows that on the 55 Likert items, with 1 as *strongly disagree* and 4 as *strongly agree*, both groups average near *agree*, with the *Building Blocks* group averaging 3.0 (*SD* = .45) and the comparison group, 2.8 (*SD* = .63). Similarly, there were few notable differences on the subscale scores. The comparison teachers scored somewhat higher on using management strategies to enhance the quality of lessons (two items), conducting the activity as written (two items), encouraging mathematical reflection, and using scaffolding activities. The *Building Blocks* teachers scored somewhat higher on staying on schedule, sending activities home, completing activities with all children, asking children to share and justify ideas, allowing children to select center activities, being actively involved, promoting effort, monitoring the activities, and accessing software records.

The repeated-measures ANOVA computed on the Fidelity *T* score was not significant for time (fall, winter, spring), $F(2, 38) = .33$, $p = .73$, $MSE = 47.26$); treatment $F(1, 19) = .07$, $p = .80$, $MSE = 198.38$; or Treatment × Time interaction, $F(2, 38) = .30$, $p = .74$, $MSE = 47.26$. There also were no significant interactions of treatment by program type (Head Start vs. state funded), $F(1, 17) = .2.03$, $p = .17$, $MSE = 165.42$, or of Time × Treatment × Program Type, $F(2, 34) = .94$, $p = .40$, $MSE = 49.17$. Finally, the Fidelity scores correlated positively, but not significantly, with children's gain scores, $r = .19$, $p = .40$. Thus, total fidelity scores were acceptably positive, but there is no evidence that they changed over time, were different in the two intervention groups, or interacted with program type.

Quality of the Mathematics Environment and Teaching (COEMET)

Table 4 presents the means and standards deviations for the *T* score and means for the descriptive items and subtests for the COEMET. To measure the relative effects of the three treatment groups on the quality of the mathematics environment and teaching, a repeated-measures ANOVA was computed on the *T* score. This analysis yielded a significant treatment effect, $F(2, 32) = 6.22$, $p = .005$, $MSE = 150.10$, partial eta squared = .28 (the proportion of total variation attributable to the factor, partialing out other factors from the total nonerror variation; Pierce, Block, & Aguinis, 2004). However, there was no significant effect for time, $F(2, 64) = 3.10$, $p = .05$, $MSE = 64.94$, or for the Treatment × Time interaction, $F(4, 64) = .27$, $p = .90$, $MSE = 64.94$. There also were no significant interactions of Treatment × Program Type, $F(2, 29) = .58$, $p = .57$, $MSE = 147.61$; Time × Treatment × Program Type, $F(4, 58) = .65$, $p = .63$, $MSE = 63.80$); Treatment × Class SES, $F(1, 30) = 1.27$, $p = .27$, $MSE = 151.88$; or Time × Treatment × Class SES, $F(2, 60) = .1.94$, $p = .15$, $MSE = 57.64$.

The *Building Blocks* group had the highest scores, followed by the comparison group and then the control group, but only the comparison between the *Building Blocks* and control groups was significant, Scheffé post hoc $p = .001$, with an effect size of 1.25 (*Building Blocks* vs. comparison, $p = .06$, $ES = .92$; comparison vs. control, $p = .38$, $ES = .44$). In addition, the COEMET accounted for a significant amount of the variance in children's gain scores, $r = .49$, $p = .003$. The *Building Blocks* group was higher than the comparison and control groups on most of the items in the specific mathematics activities category, with a smaller difference on encouraging mathematical reflection (most of which had high item difficulties, indicating they were

Table 5
Means and Standards Deviations for the Early Mathematics Assessment

		Control		Comparison		*Building Blocks*	
	Max.	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
T score mean		45.47	53.22	41.30	53.77	43.02	59.39
SD		7.53	8.38	5.03	6.53	7.88	7.46
Raw subtest scores							
Number							
Verbal counting	2	1.06	1.46	0.87	1.41	0.79	1.53[a]
Recognition number, subitizing	7	2.61	3.28	1.66	2.47	2.08	3.26[a]
Object counting and strategies	25	7.24	11.03	3.59	10.33	5.06	12.61[a]
Comparing number, sequencing	28	6.14	11.74	3.44	8.78	4.66	12.31[a,b]
Composition of number	4	0.49	0.50	0.27	0.28	0.20	0.64[a,b]
Arithmetic	8	0.58	1.14	0.36	0.47	0.44	1.06[b]
Geometry, measurement, and patterning							
Shape identification	96	81.52	83.09	74.60	86.80	74.88	87.26[a]
Comparing shape	9	4.84	5.44	4.43	5.04	4.26	5.46
Representing shape	4	1.25	1.73	0.64	1.67	0.57	1.88[a]
Composing shape	10	2.70	4.06	1.47	4.35	1.86	5.57[a,c]
Transformations	1	0.28	0.33	0.25	0.25	0.24	0.43
Measurement	5	0.90	1.02	0.73	0.71	0.85	1.02
Patterning	8	2.00	3.32	0.88	3.15	1.26	3.44[a]

Note. Data from children who completed all components of both the pre- and posttest and thus were included in the hierarchical linear model analyses; n = control, 101; comparison, 51; and *Building Blocks*, 101. Max. = Maximum raw scores for each subtest. Shape-identification items, four items that showed the same 26 shapes and asked children to find all exemplars of a shape category (e.g., squares), were weighted in the Rasch analysis.
[a]Significantly higher than control group using the reduced hierarchical linear model, $p < .05$.
[b]Significantly higher than comparison group, $p < .05$.
[c]$p < .10$.

relatively more difficult to achieve for this population). The *Building Blocks* group employed a greater number of different activities than the other two groups. The number of computers running mathematics activities was highest in the *Building Blocks*, then comparison, then control group. Both intervention groups were slightly higher in the percentage of time children used the computers and in the percentage of time the teachers were in the room. As an exploratory analysis, stepwise regression was performed to test a hypothesis that items of greater Rasch difficulty would distinguish more (from less) effective teachers. Three individual items predicted children gain scores significantly, $p < .0001$. The percentage of time the teacher was

Table 6
Hierarchical Linear Models

Main Model[a]				
Fixed Effects	Coefficient	*SE*	*t*	*p*
Intercept	11.77	0.41	28.94	000+
Program type	–1.10	1.12	–0.99	.33
Class socioeconomic status (SES)	0.11	1.04	0.11	.91
Building Blocks (*BB*)	8.47	0.92	9.20	000+
Comparison	4.79	1.28	3.74	000+
Interaction program Type w/*BB*	–0.84	2.47	–0.34	.74
Interaction program type w/comparison	–1.66	3.12	–0.53	.60
Interaction class SES w/*BB*	0.24	2.03	0.12	.90
Random Effect	*SD*	Var	χ^2	*p*
Intercept	1.56	2.42	47.55	.009
Level 1	4.80	23.05		
***BB* Group Compared to the Comparison Group**				
Fixed Effects	Coefficient	*SE*	*t*	*p*
Intercept	14.75	0.55	26.87	000+
BB vs. comparison	3.55	1.16	3.05	.007
Random Effect	*SD*	Var	χ^2	*p*
Intercept	1.67	2.82	34.43	.016
Level 1	5.02	25.21		

Note. Var = variance component, the between-classroom variance.
[a]All *df* = 27.
[b]All *df* = 19.

actively engaged in activities was of moderate difficulty. The degree the teacher built on and elaborated children's mathematical ideas and strategies was the highest difficulty item. The degree the teacher facilitated children's responding was a moderate difficulty item. Recall that these items are influenced by the number of activities conducted. Therefore, there is some, but only moderate, support for the notion that items of greater Rasch difficulty would be particularly useful in identifying the most effective teachers.

Children's Mathematics Achievement

The third research question was, What are the effects of the *Building Blocks* curriculum, as implemented under diverse contexts, on the mathematics achievement of preschoolers?

Total test. Table 5 presents the means and standards deviations for the *T* score and means for the descriptive items and subtests for the EMA for all children for whom full data were collected for both pretest and posttest. Table 6 presents the HLM analyses on these data. We analyzed main effects, first, by comparing the two intervention groups to the control group in the main model and, second, by running a reduced model comparing the *Building Blocks* group to the comparison group. Both intervention groups significantly outperformed the control group (first model, *Building Blocks*, $p = 000^{+}$; comparison, $p = 000^{+}$), and the *Building Blocks* group significantly outperformed the comparison group (second model, *Building Blocks* vs. comparison, $p = .007$). There were no main effects for class SES or program type.

The effect size for the *Building Blocks* group compared to the control group was 1.07 (the standardized mean difference *ES* calculated from child-level pre- and posttest scores and the pooled pretest standard deviation was a comparable 1.09), and for the *Building Blocks* group compared to the comparison group, .47 (compared to .54). The effect size for the comparison group compared to the control group was .64 (compared to .60).

Table 6 also presents the interactions between the intervention groups and the Level 2 predictors. There were no significant interactions between program type and either intervention group nor between the *Building Blocks* group and class SES. Thus, there was no evidence that the interventions were more effective with either program type and no evidence that the *Building Blocks* curriculum was differentially effective in classes serving low- or mixed-income families. (Interactions between the Level 1 variables of child-level SES and gender and Level 2 treatment groups, entered in an initial model, revealed no significant effects. Due to the exploratory nature of these Level 1 interactions, these are not included in the model presented in Table 6.)

Finally, the random effect intercept reveals significant unexplained variance. A following section discusses the mediational hypothesis, which was tested to see if this variance can be partially explained as a function of differences in observations of the quality and quantity of mathematics teaching and environment.

Specific topics and strategies. The descriptive statistics in Table 5 suggest that the three treatment conditions differentially affected certain categories of mathematics knowledge and competencies. Consistent with the *T* score results, the *Building Blocks* group scored higher than the comparison group, which scored higher than the control group, on the single item on verbal counting. Table 7 provides descriptive statistics for individual items. (The order of these items is as they occurred on the EMA, according to increasing Rasch difficulty, but the numbering is consecutive in Table 7 for ease of referral. Recall that there was a stop rule; thus, not all children were administered every item.) There were few differences in the types of errors that children made (see Table 7, Item 1). The main differences were simply that the *Building Blocks* group counted farther without errors than the other two

(text continues on p. 482)

Table 7
Percentages for Solution Strategies and Error Types on the Early Mathematics Assessment

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
<u>Number</u>						
Verbal counting						
1. Count verbally						
Counted through decades, logical invented patterns	1.0	0.0	0.0	0.0	0.0	0.0
Skipped one number then resumed sequence	17.8	21.8	16.7	27.1	23.8	22.8
Omitted more than one number word	15.8	9.9	27.1	10.4	13.9	5.9
Repeated some number words	3.0	0.0	2.1	6.3	3.0	4.0
No mistakes in the number sequence achieved	0.0	44.6	0.0	39.6	0.0	52.5
Other	52.5	3.0	41.7	2.1	44.6	0.0
No response	9.9	20.8	12.5	14.6	14.9	14.9
Object counting and strategies						
2. Count five objects, arranged in a line						
Did not point or touch (and correct)	6.0	11.8	6.3	6.4	7.0	16.9
Correctly pointed or touched	80.2	77.2	52.1	91.5	71.0	72.3
Pointed at some more than once	2.0	0.0	8.3	2.1	3.0	1.0
Skipped some	1.0	2.0	10.4	0.0	2.0	0.0
Skim or flurry error	3.0	4.0	0.0	0.0	7.0	1.0
Omitted or repeated saying some number words	3.0	0.0	8.3	0.0	1.0	0.0
Repeated saying some number words	0.0	0.0	2.1	0.0	0.0	0.0
Other	3.0	5.0	8.3	0.0	8.0	6.8

(continued)

Table 7
(continued)

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
No response	2.0	0.0	2.1	0.0	1.0	1.0
Item not administered	0.0	0.0	2.1	0.0	0.0	1.0
3. Identify another's mistake in object counting						
Incorrect	21.8	15.8	18.8	4.2	20.8	6.9
Correct, showed mistake	43.6	62.4	27.1	64.6	27.7	71.3
Indicated mistake but did not explain it	9.9	13.9	6.3	20.8	11.9	18.8
No response	0.0	0.0	2.1	0.0	2.0	0.0
Item not administered	24.8	7.9	45.8	10.4	37.6	3.0
4. Shown six objects, produce six						
a. Duplicated the spatial arrangement	0.0	2.9	1.8	2.0	0.9	4.0
b. Counted your six, then the child's six	8.3	28.4	0.0	23.5	3.6	36.6
c. Both a and b	0.9	0.0	1.8	2.0	0.0	3.0
d. Stated *six*, no verbal counting	0.9	2.0	0.0	2.0	0.9	1.0
e. Other, incorrect	5.6	2.0	12.5	15.7	5.4	5.0
f. Other, correct	2.8	1.0	1.8	5.9	0.9	6.9
g. No visible strategy, incorrect	8.9	32.2	25.0	37.3	32.4	33.7
h. No visible strategy, correct	38.3	19.6	0.0	0.0	3.6	6.9
i. No response, incorrect	1.9	0.0	5.4	0.0	0.0	0.0
j. Item not administered	32.4	11.9	51.8	11.8	52.3	3.0

(continued)

Table 7
(continued)

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
5. Count scrambled arrangement of 15 objects						
Reading order	6.0	2.0	0.0	4.3	3.0	1.0
Any other consistent movement	13.0	15.0	0.0	12.8	11.0	5.0
Spiraling in	8.0	3.0	2.1	0.0	0.0	0.0
Meandering path	17.0	19.0	12.8	21.3	8.0	27.0
Moved to a new pile	14.0	41.0	12.8	38.3	12.0	41.0
Reordered before counting	3.0	3.0	6.4	6.4	2.0	10.0
Random or no response	3.0	3.0	4.3	2.1	7.0	11.0
Item not administered	36.0	14.0	61.7	14.9	57.0	5.0
6. Count unorganized arrangement of 30 objects						
Reading order	14.9	32.7	8.3	35.4	15.8	30.3
Other systematic spatial strategy	21.8	35.6	22.9	33.3	16.8	51.5
Spiraling in	13.9	1.0	2.1	0.0	2.0	0.0
Meandering path	7.9	5.0	2.1	10.4	2.0	2.0
Random or no response	3.0	8.9	0.0	2.1	3.0	10.1
Item not administered	38.6	16.8	64.6	18.8	60.4	6.1
7. Count to 10, starting at 4						
Incorrect	48.5	29.7	50.0	31.3	45.5	19.8
Counted up from 1 with no errors	16.8	14.9	12.5	27.1	11.9	18.8
Began with 4 or 5, reached 10 but makes error	3.0	3.0	6.3	0.0	5.9	3.0
Immediately began at 4 and counted to 10	19.8	46.5	2.1	27.1	6.9	46.5
No response	6.9	3.0	4.2	12.5	9.9	10.9
Item not administered	5.0	3.0	25.0	2.1	19.8	1.0

(continued)

Table 7
(continued)

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Comparing number and sequencing						
8. Compare number of objects (3, 3, different arrangements)						
Incorrect	37.6	35.6	43.8	31.3	46.5	20.8
Correct	61.4	64.4	50.0	68.8	50.5	79.2
No response	1.0	0.0	6.3	0.0	3.0	0.0
9. Compare number of objects (3, 4, same area)						
Scrambled	6.9	5.9	14.6	4.2	6.9	3.0
Line	63.4	81.2	39.6	79.2	57.4	90.1
Incorrectly decided that they are the same	4.0	0.0	0.0	0.0	9.9	1.0
Other	0.0	1.0	6.3	16.7	1.0	2.0
No response	9.9	11.9	4.2	0.0	4.0	4.0
Item not administered	15.8	0.0	35.4	0.0	20.8	0.0
10. Comparing number of objects (4 large, 5 small)						
Referred to numbers without visible counting	51.8	2.9	1.7	3.9	11.7	8.1
Counted	6.0	67.6	42.4	70.6	45.9	65.7
Matched	0.0	0.0	0.0	3.9	3.6	0.0
Other, incorrect	0.9	2.9	10.2	7.8	1.8	2.0
Other, correct	0.0	2.0	0.0	0.0	0.9	4.0
No visible strategy, incorrect	21.1	6.9	22.0	3.9	20.7	4.0
No visible strategy, correct	9.2	13.7	10.2	5.9	9.0	14.1
No visible strategy, no response	4.6	3.9	5.1	3.9	1.8	2.0
Item not administered	6.4	0.0	8.5	0.0	4.5	0.0

(continued)

Table 7
(continued)

Item/Response	Control		Comparison		*Building Blocks*	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
11. Comparing number of objects (9 large, 11 small)						
Referred to numbers without visible counting	0.0	0.0	0.0	0.0	3.6	0.0
Counted	14.7	56.9	5.3	54.9	9.0	61.6
Matched	0.9	1.0	1.8	2.0	9.9	0.0
Other, incorrect	0.0	1.0	0.0	2.0	0.0	1.0
Other, correct	0.0	1.0	0.0	2.0	0.0	1.0
No visible strategy, incorrect	0.0	22.5	3.5	29.4	10.9	22.2
No visible strategy, correct	2.8	2.0	0.0	0.0	0.0	4.0
No response, incorrect	3.7	2.0	5.3	0.0	15.4	1.0
No response, correct	0.0	0.0	1.8	0.0	0.0	0.0
Item not administered	78.0	13.7	82.5	9.8	51.2	9.1
12. Which number is closer to 7: 3 or 4?						
Kept track while counting by ones	0.0	2.9	0.0	0.0	0.0	4.0
Said 3 is before 4 or 4 is after 3	0.9	4.9	0.0	0.0	0.0	4.0
Other, incorrect	0.9	0.0	0.0	2.0	0.0	4.0
Other, correct	2.8	15.8	0.0	5.9	0.9	10.9
No visible strategy, incorrect	3.7	4.9	0.0	5.9	0.0	5.9
No visible strategy, correct	3.7	2.0	1.7	2.0	0.9	7.9
No response, incorrect	0.0	12.7	0.0	5.9	0.9	5.0
No response, correct	0.9	5.9	0.0	5.9	0.9	7.9
No response, no response	1.9	1.0	0.0	2.0	0.0	3.0
Item not administered	85.2	50.0	98.4	70.6	96.5	47.6

(continued)

Table 7
(continued)

Item/Response	Control		Comparison		*Building Blocks*	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
13. Which number is closer to 6: 9 or 4?						
Kept track while counting by ones	0.0	3.0	0.0	0.0	0.0	2.0
Other, incorrect	0.9	0.0	0.0	0.0	0.9	6.1
Other, correct	1.8	12.0	0.0	2.0	0.0	11.1
No visible strategy, incorrect	1.8	4.0	0.0	5.9	0.9	8.1
No visible strategy, correct	5.5	2.0	1.7	3.9	0.0	2.0
No response, incorrect	0.9	13.0	0.0	9.8	0.0	12.1
No response, correct	0.9	9.0	0.0	0.0	0.9	5.1
No response, no response	0.9	0.0	0.0	0.0	0.0	2.0
Item not administered	87.3	57.0	98.4	78.4	97.4	51.5
Arithmetic						
14. Addition (2 + 1)						
Uses objects	3.7	3.9	0.0	3.9	0.9	7.0
Added on	2.8	0.0	1.7	0.0	0.9	4.0
Kept track while counting on	0.0	1.0	0.0	2.0	0.9	1.0
Derived combination	0.0	0.0	1.7	0.0	0.0	0.0
Combination	0.0	0.0	0.0	2.0	0.9	1.0
Estimation or guess	2.8	0.0	13.6	0.0	13.5	1.0
Other, incorrect	18.3	6.9	20.3	19.6	9.0	19.0
Other, correct	0.0	3.9	3.4	3.9	0.9	3.0
Other, no response	0.0	1.0	3.4	2.0	0.0	0.0
No visible strategy, incorrect	50.5	56.9	37.3	47.1	45.0	29.0
No visible strategy, correct	13.8	25.5	5.1	9.8	9.0	22.0

(continued)

Table 7
(continued)

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
No response, incorrect	0.0	0.0	6.8	0.0	3.6	0.0
No response, no response	8.3	1.0	6.8	9.8	14.4	13.0
Item not administered	0.0	0.0	0.0	0.0	0.9	0.0
15. Addition (3 + 2)						
Objects	1.8	5.9	1.7	2.0	1.8	7.1
Added on	0.0	1.0	1.7	0.0	0.9	5.1
Kept track while counting on	1.8	2.0	1.7	2.0	0.0	1.0
Derived combination	0.9	1.0	0.0	0.0	0.0	0.0
Combination	0.9	0.0	0.0	0.0	0.0	0.0
Estimation or guess	1.8	0.0	10.3	0.0	14.4	1.0
Other, incorrect	11.0	3.9	24.1	13.7	6.3	14.1
Other, correct	0.9	2.0	0.0	3.9	0.0	1.0
No visible strategy, incorrect	60.6	49.0	43.1	60.8	54.1	43.4
No visible strategy, correct	11.0	28.4	10.3	7.8	7.2	18.2
No visible strategy, no response	0.0	0.0	0.0	2.0	0.0	0.0
No response, incorrect	0.0	0.0	1.7	0.0	2.7	0.0
No response, no response	9.2	6.9	5.2	7.8	11.7	8.1
Strategy missing						
Item not administered	0.0	0.0	0.0	0.0	0.9	1.0
<u>Geometry and Patterning</u>						
Representing shape						
16. Construct triangle with straws (accuracy)						
Not at all correct	36.63	23.76	60.78	23.53	59.41	14.85
"Partially correct" (basic spatial arrangement)	45.54	52.48	25.49	58.82	24.75	53.47
Completely correct	7.92	20.79	3.92	15.69	1.98	24.75

(continued)

Table 7
(continued)

Item/Response	Control Pretest	Control Posttest	Comparison Pretest	Comparison Posttest	*Building Blocks* Pretest	*Building Blocks* Posttest
No response	8.91	2.97	9.80	1.96	12.87	6.93
Item not administered	0.99	0.00	0.00	0.00	0.99	0.00
17. Construct triangle with straws (angles)						
Left gaps	21.78	19.80	3.92	17.65	6.93	13.86
Ends of lines stick out	15.84	17.82	11.76	25.49	5.94	23.76
Both gaps and lines	9.90	13.86	11.76	13.73	11.88	11.88
Correctly connected corner	7.92	20.79	1.96	17.65	2.97	28.71
No response	43.56	27.72	70.59	25.49	71.29	21.78
Item not administered	0.99	0.00	0.00	0.00	0.99	0.00
18. Construct rectangle with straws (accuracy)						
Not at all correct	36.63	20.79	58.82	25.49	48.51	14.85
"Practically correct" (basic spatial arrangement)	40.59	60.40	23.53	60.78	26.73	55.45
Completely correct	11.88	9.90	3.92	7.84	1.98	14.85
No response	9.90	8.91	13.73	5.88	21.78	14.85
Item not administered	0.99	0.00	0.00	0.00	0.99	0.00
19. Construct rectangle with straws (angles)						
Left gaps	8.91	12.87	0.00	5.88	2.97	15.84
Ends of lines stick out	8.91	20.79	9.80	11.76	4.95	16.83
Both gaps and lines	20.79	21.78	11.76	43.14	15.84	16.83
Partial shape	0.99	2.97	1.96	0.00	1.98	3.96
Correctly connected corners	13.86	11.88	3.92	7.84	2.97	16.83
No response	45.54	29.70	72.55	31.37	70.30	29.70
Item not administered	0.99	0.00	0.00	0.00	0.99	0.00

(continued)

Table 7
(continued)

Item/Response	Control Pretest	Control Posttest	Comparison Pretest	Comparison Posttest	*Building Blocks* Pretest	*Building Blocks* Posttest
20. Construct rectangle with straws (angle size)						
Incorrect on all	4.95	7.92	5.88	0.00	2.97	0.99
One to three right angles	12.87	12.87	11.76	23.53	11.88	11.88
All four right angles	33.66	48.51	9.80	35.29	10.89	47.52
One to four corners with gaps but perpendicular	1.98	0.99	3.92	9.80	1.98	9.90
No response	45.54	29.70	68.63	31.37	71.29	29.70
Item not administered	0.99	0.00	0.00	0.00	0.99	0.00
Composing shape						
21. Composing shape (accuracy)						
Placed no shapes or placed shapes but not one "fits"	13.86	3.96	39.22	0.00	43.56	0.99
> 0% but < 50% of shapes placed fit or >2 gaps	41.58	43.56	50.98	29.41	34.65	13.86
50+% fit, but one to two gaps, or shapes hang over line	28.71	27.72	7.84	27.45	16.83	16.83
100% of shapes placed fit, but one to two gaps	3.96	4.95	0.00	11.76	0.00	2.97
Covered puzzle, but one to two shapes hang over the line	1.98	0.00	0.00	3.92	0.00	2.97
Completed puzzle accurately (no gaps or "hangovers")	6.93	19.80	0.00	25.49	2.97	61.39
No response	1.98	0.00	1.96	1.96	1.98	0.99
Item not administered	0.99	0.00	0.00	0.00	0.00	0.00

(continued)

Table 7
(continued)

Item/Response	Control		Comparison		*Building Blocks*	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
22. Composing shape (accuracy)						
Turned shapes after placing on puzzle in an attempt to fit	11.88	3.96	1.96	7.84	6.93	1.98
Turned shapes into correct orientation before placing	31.68	48.51	5.88	60.78	12.87	82.18
NA (Code 16A = 0 or 1)	55.45	47.52	92.16	31.37	80.20	15.84
Item not administered	0.99	0.00	0.00	0.00	0.00	0.00
23. Composing shape (accuracy)						
Seemingly trial and error	7.92	7.92	1.96	11.76	2.97	3.96
Appeared to search, then found and placed it.	33.66	44.55	5.88	56.86	16.83	80.20
NA (Code 16A = 0 or 1)	57.43	47.52	92.16	31.37	80.20	15.84
Item not administered	0.99	0.00	0.00	0.00	0.00	0.00
24. Composing shape (accuracy)						
Hesitant or not systematic	2.97	4.95	1.96	3.92	3.96	2.97
Solved the puzzle systematically but may be “halting”	24.75	34.65	3.92	41.18	15.84	39.60
Solved the puzzle systematically, immediately, confidently	13.86	12.87	1.96	23.53	1.98	41.58
NA (Code 16A = 0 or 1)	57.43	47.52	92.16	31.37	78.22	15.84
Item not administered	0.99	0.00	0.00	0.00	0.00	0.00

(continued)

Table 7
(continued)

Item/Response	Control Pretest	Control Posttest	Comparison Pretest	Comparison Posttest	*Building Blocks* Pretest	*Building Blocks* Posttest
Patterning						
25. Patterning (A-B-A-B-A-B)						
Incorrect	29.70	21.78	64.71	29.41	52.48	17.82
Began pattern or had some of core but had error	17.82	8.91	11.76	13.73	11.88	13.86
Correctly completed one or more full units but had extra	8.91	6.93	0.00	9.80	1.98	7.92
Correctly completed one or more full units	39.60	62.38	13.73	47.06	26.73	58.42
No response	2.97	0.00	9.80	0.00	5.94	0.99
Item not administered	0.99	0.00	0.00	0.00	0.99	0.99
26. Patterning (A-B-B-A-B-B-A-B-B)						
Incorrect	72.28	65.35	76.47	72.55	86.14	64.36
Began pattern or had some of core but had error	15.84	15.84	11.76	15.69	7.92	14.85
Completed one or more full units but had incomplete unit	0.00	4.95	0.00	5.88	0.00	2.97
Correctly completed one or more full units	5.94	12.87	0.00	3.92	0.00	10.89
No response	4.95	0.99	11.76	1.96	4.95	5.94
Item not administered	0.99	0.00	0.00	0.00	0.99	0.99

groups, and the comparison group counted farther with no errors than the control group.

The *Building Blocks* group scored higher than both the comparison and control groups on recognition of number and subitizing (Table 5). Relative gains on recognition of number were primarily in improving quick and accurate recognition of small numbers (e.g., 2 and 4).

Both intervention groups scored higher than the control group, with little difference between them, on object counting and verbal counting strategies, comparing number, and sequencing (Table 5). On object counting, the most consistent relative gains were on simple object counting and production ("Give me six . . .") tasks as well as some sophisticated counting strategy tasks (e.g., "Here are six pennies. There are three more under this cloth. How many are there in all?"). By the posttest, 95% of all children used effective counting strategies for counting five objects in a line. The comparison group used such strategies less at pretest than the other two groups and thus showed more increase in the behavior of correctly touching or pointing to each object. The *Building Blocks* group, more than the other groups, decreased in its frequency of skim or flurry errors and increased visual-only strategies in maintaining correspondences (Table 7, Item 2). Both intervention groups made more progress than the control group in identifying errors in counting, with the *Building Blocks* group more likely to describe the error made and how to correct it (Item 3). On a task (Item 4) in which six boxes were shown in a random arrangement and the child was asked to put the same number of boxes in his or her shopping cart, both intervention groups, but especially the *Building Blocks* group, increased the use of the strategy of counting the target group, then counting out an identical number. The *Building Blocks* group also increased the frequency of accurate mental strategies. Several items asked children to count large, unordered collections. Children's most preferred strategy (about 40% of the children at posttest; see Item 5) was to move the items when that was possible, with following a meandering path (19% to 27%) about equal in frequency to more structured spatial-counting strategies (e.g., reading order, spiraling in) in preference. Both intervention groups increased systematic strategies more than the control group. When moving the objects was not possible (Item 6), about 70% to 80% of the children followed reading order or another consistent spatial strategy, with only a moderate increase of the two intervention groups relative to the control group on systematic strategies. The comparison group increased the use of "reading order," whereas the *Building Blocks* group showed a larger increase in the use of other systematic strategies.

Substantially more children in both intervention groups, compared to the control group, reached items on the EMA that required children to count on or back from numbers other than 1. The *Building Blocks* group made somewhat more progress than the other two groups in verbally counting on starting from 4 (Item 7).

Gains for the intervention groups were most pronounced on comparing small sets (less than five; Items 8 and 9) and ordering numerals 1 to 5. On some items (10 and 11), there was a trend for both intervention groups, compared to the control group, to engage in more overt counting and less matching and for the *Building Blocks* group to show increased use of mental strategies. The *Building Blocks* group was slightly less likely to use matching or subitizing on the posttest than the other groups. Other items had low responses, but the *Building Blocks* group showed a similar increase in the use of mental strategies (Items 12 and 13).

There were no consistent gains on ordering numbers, identifying the smaller of two sets or numbers, or identifying which of two numbers was closer to a third number, although the *Building Blocks* group gained more than either other group (Table 5).

The *Building Blocks* group scored higher than both the comparison and control groups on arithmetic (Table 5). This comparison was highest on additive complement items (instant recognition of parts and wholes), on which the *Building Blocks* group gained (the strategy is emphasized in that curriculum) but the other groups declined (perhaps due to their emphasis on counting-based arithmetic). The *Building Blocks* group slightly increased the frequency of using objects and adding on compared to both the comparison and control groups (Table 7, Items 14 and 15). Both intervention groups decreased their use of guessing and other uncategorizable strategies that lead to incorrect responses.

Turning to geometry, both intervention groups scored higher than the control group, with little difference between them, on identifying shapes and representing shapes (Table 5). Descriptions in Table 7 (Items 16 to 20) indicate gains of both the intervention and comparison groups relative to the control group in producing "partially correct" representations (basic spatial configurations in building a shape with straws). The *Building Blocks* group also increased in the frequency of completely correct constructions more than the other two groups. On the shape-identification items, children gained on most of the individual shapes, with the greatest gains on prototypes and rotated variants of the class for squares and triangles and for these and particular distractors for rectangles (e.g., avoiding choosing a parallelogram) and rhombuses.

The *Building Blocks* group scored higher than both the comparison and control groups on comparing shape (Table 5). Children made greater gains increasing their matches of congruent shapes (requiring slides, flips, or turns) than on decreasing erroneous pairing of noncongruent shapes.

Consistent with the *T* score results, the *Building Blocks* group scored higher than the comparison group, which scored higher than the control group, on shape composition (Table 5). Again, the comparison group's greater gains than the control group resulted from an increase in partially correct solutions. The *Building Blocks* group increased more than the other two groups in completely correct solutions (Table 7, Item 21). Similarly, the *Building Blocks* group increased substantially more than the other groups in using more sophisticated strategies, such as rotating shapes into the correct orientation before placing them on the puzzle (Item 22), searching for specific shapes with intentionality (Item 23), and, in general, solving the puzzle systematically, immediately, and confidently (Item 24). A caveat is that most of the items on which children gained were similar to activities in the intervention curricula.

Although differences were small, the control group made the greatest gains on the measurement subtest (Table 5). Finally, on the patterning items, both intervention groups scored higher than the control group, with little difference between them. The intervention groups made gains on copying an A-B-A-B, but much smaller gains on an A-B-B-A-B-B, pattern. Relative increases in frequency were made in partial-credit behaviors as well as in providing a completely correct solution but, again, more for A-B-A-B (Table 7, Item 25) than for A-B-B-A-B-B (Item 26) patterns. Other items were mixed. Both groups gained more than the control group on completing patterns and identifying the core unit of patterns, with the *Building Blocks* group making more consistent gains; however, differences were small.

Test of Mediation

The fourth question was whether an observable change in the quality of the preschool mathematics environment and teaching mediated the effects on mathematics achievement. That is, does the COEMET function as a mediator between treatment and child outcomes? The equations of the multiple-regression approach (Baron & Kenny, 1986) were significant: Equation 1, regressing the COEMET on treatment, $R = .32$, $R^2 = .10$, $df = 1, 43$, $MSE = 59.86$, $p < .05$, $\beta = .32$; Equation 2, achievement gain on treatment, $R = .57$, $R^2 = .33$, $df = 1, 43$, $MSE = 12.87$, $p < .0001$, $\beta = .57$; and Equation 3, achievement gain on COEMET and treatment, for the COEMET mediator, $R = .50$, R^2 change $= .25$, $df = 1, 43$, $MSE = 14.40$, $p < .0001$, $\beta = .50$, and for treatment, $R = .66$, R^2 change $=.19$, $df = 1, 42$, $MSE = 11.00$, $p < .01$, $\beta = .46$. Because the independent variable (treatment) still accounted for a significant amount of the variance in Equation 3, there was not strong evidence for mediation. However, the amount of variance attributable to treatment group was reduced substantially to .19 from .33, providing evidence for partial mediation.

Discussion

The main purpose of this research was to measure the effectiveness of a preschool mathematics program based on a comprehensive model of developing research-based curriculum in larger contexts with teachers and students of diverse backgrounds (the final phase of the CRF model, Summative Research: Large Scale). The first research question was, Can *Building Blocks* be implemented with high fidelity, and does the measure of fidelity predict achievement gains? Teachers implemented both intervention curricula, *Building Blocks*, and the comparison mathematic curriculum with acceptable fidelity across the three measurement periods and no evidence of change across those periods (fall, winter, spring). The average of the *Building Blocks* teachers' Fidelity scores corresponded exactly with *agree* on the Likert scale, and the comparison teachers' scores were only slightly lower. Although small, the differences on individual items on this scale were consistent with differences in the curricula. For example, the comparison curriculum's print materials emphasize carefully managed activities, including prescribed scaffolding strategies. The *Building Blocks* curriculum emphasizes continuous active involvement of teachers to engage all children in trying hard to solve problems, to encourage them to communicate and justify solutions to these problems, and to monitor children's progression through learning trajectories in hands-on and computer activities, adapting activities as needed. Fidelity scores positively correlated with children's gains in achievement, but this did not reach statistical significance, possibly because there was insufficient variance in the scores. Future research with greater numbers of teachers and greater variance in scores should ascertain whether significant relations emerge under these conditions.

In summary, results indicate that research-based mathematics preschool curricula can be implemented with good fidelity from the beginning of the year, at least if teachers are provided training and support to the extent involved in the present study. This involved 34 hours of focused group work and about 16 hours of in-class coaching.

The second research question was, Does *Building Blocks* have substantial positive effects on the quality of preschool mathematics environments and teaching? The *Building Blocks* curriculum had a significant positive effect on the quality of the preschool mathematics observed in the classroom environment and teaching relative to the control condition. *Building Blocks* teachers provided a greater number of different activities and more computer activities and interacted with children in the context of those activities more than the other groups. Although the effect relative to the comparison curriculum did not reach statistical significance, $p = .06$, the small number of teachers (e.g., 7 in the comparison group, resulting in low statistical power) and large effect size, $d = .92$, suggest that this comparison be reinvestigated. Future studies may wish to supplement the instrument used in this study, which was intentionally designed not to favor any particular curriculum and to investigate specific features such as the use of learning trajectories.

The third research question was, What are the effects of the *Building Blocks* curriculum, as implemented under diverse contexts, on the mathematics achievement of preschoolers? The achievement gains of the *Building Blocks* group compared to the control group exceeded those considered large (Cohen, 1977, uses .8 as the benchmark for large effects; Rosenthal & Rosnow, 1984, use .5). This confirms the first summary evaluation (Clements & Sarama, 2007c) and is particularly significant, considering that the present study involved a greater number of, and more diverse, classrooms and that research staff were not present in classrooms on a daily basis, as they were in the previous study. The comparison curriculum also produced significantly more achievement gain than the control condition. Thus, preschool math curricula can lead to gains. What advantages are there to the unique features of the *Building Blocks* curriculum?

The moderate-sized effect of the *Building Blocks* group curriculum above that of the intensive mathematics curriculum of the comparison group gives additional support to the hypothesis that an educational program based on a comprehensive model of developing research-based curriculum, the CRF (Clements, 2007), can be uniquely effective. Consistent with that model, *Building Blocks* is based on a core of interwoven learning trajectories. Differences on the observational instruments show that *Building Blocks* teachers were more likely than comparison, as well as control, teachers to monitor and be actively involved with activities, using formative assessment based on their knowledge of children's developmental progressions and a greater number of activities linked to children's developmental level, all indicating their use of the learning trajectories.

There was no evidence that *Building Blocks* was differentially effective with Head Start versus state-funded preschools or with classes serving low- versus mixed-income families. Exploratory analyses also revealed no evidence that the *Building Blocks* curriculum was differentially effective for children from low- or middle-SES households or of different gender. Therefore, there is no reason to suspect that the CRF is differentially effective for any these subpopulations.

Both intervention groups outperformed the control group on verbal counting strategies, sequencing, identifying shapes, representing shapes, and patterning. All three curricula cover topics in approximately the same proportion, but individual teachers may emphasize topics familiar to them. However, similar relative gains in some of the topics emphasized in all treatment groups, such as object counting and comparing number, suggest that research-based activities in the intervention groups may have been particularly effective. The *Building Blocks* curriculum made the most substantial gains relative to both other groups in verbal counting, recognition of number and subitizing, comparison of shape, and shape composition. Examination of children's behaviors on individual items suggests that these children were more accurate and increased the use of more sophisticated mental strategies. These results are similar to those of the previous summary research on the *Building Blocks* curriculum (Clements & Sarama, 2007c). Given that both intervention curricula cover these topics in similar proportions, the learning trajectories structure of the *Building Blocks* curriculum is the likely reason for the greater gains in these topics.

The fourth question was whether an observable change in the quality of the preschool mathematics environment and teaching mediated the effects on mathematics achievement. The statistical model supported a partial mediation hypothesis. Thus, the *Building Blocks* curriculum and training increased the observable mathematics in the environment and the quality of teachers' pedagogical strategies, and results suggest that this increase partially accounted for children's achievement gains. However, the direct effect of providing the curricula, as well as training and coaching in its use, was substantial, beyond the effect these had on observable features of the environment and teaching. Inducing teachers to enact the basic core of effective activities may alone increase children's learning.

As a summary, results support the hypothesis that the research-based *Building Blocks* curriculum can be implemented with acceptable fidelity in multiple, diverse classrooms. Further, they support the hypotheses that such a high-quality implementation can increase both the quality of the classroom environment and teaching and preschoolers' mathematics achievement, even when compared to another intensive preschool mathematics curriculum. Finally, there is some support for the mediational hypothesis that the increase in quality and quantity of mathematics in the classroom environment and teaching accounts for the increase in preschoolers' achievement. However, other effects, most reasonably, the direct effects of engaging in mathematics activities, also accounted for these achievement gains. Research with larger numbers of teachers that also investigates sustainability of the intervention should investigate these questions in more detail.

Implications and Caveats

The results have implications in four areas: the effects of curriculum materials, the amount of professional development, the usefulness of the CRF, and the efficacy of a specific component of that framework, the learning trajectories construct.

Evaluation of Curriculum Materials

This study adds much-needed research on the effects of curriculum materials (e.g., NRC, 2004; Senk & Thompson, 2003), as it provides evidence for the effectiveness of a curriculum built on comprehensive research-based principles. Considered along with previously reported research (Clements & Sarama, 2004a, 2007c; Sarama, 2004; Sarama & Clements, 2002), the evidence includes all components of high-quality evaluations of the effectiveness of a mathematics curriculum (NRC, 2004). The present study's use of a randomized-trials design and two counterfactual groups, one taught with a different intensive mathematics curriculum, is noteworthy in meeting rigorous criteria for curriculum evaluation. A caveat is that the same research team developed the assessment of child outcomes and the *Building Blocks* curriculum. However, the assessment was developed and reviewed independently, and its topics corresponded equally well with the comparison curriculum. Indeed, the comparison curriculum addresses every topic on the assessment, including more extensive coverage than the *Building Blocks* curriculum of some topics, such as patterning and measurement. Finally, the control condition showed no advantage in any topic, including those that their curricula emphasized, such as numbers and simple arithmetic.

Results also support previous studies showing that organized experiences result in greater mathematics knowledge upon entry into kindergarten (Bodovski, 2006; Bowman et al., 2001; Shonkoff & Phillips, 2000) and that focused early mathematical interventions help young children develop a foundation of informal mathematics knowledge (Clements, 1984), including children living in poverty (Campbell & Silver, 1999; Fuson, Smith, & Lo Cicero, 1997; Griffin, 2004; Griffin, Case, & Capodilupo, 1995; Ramey & Ramey, 1998). They extend this research by suggesting that mathematics curricula (both the *Building Blocks* and comparison curricula) can increase knowledge of multiple mathematical concepts and skills beyond number. Consider the differences between low-SES children and their peers upon entry to kindergarten: .55 standard deviations below middle-SES children and 1.24 standard deviations below high-SES children (Lee & Burkam, 2002). The substantial positive effects of the *Building Blocks* curriculum compared to the control condition suggests it could help to ameliorate the problem posed by low-SES children's lack of mathematics achievement upon entrance to kindergarten. There was no evidence that positive effects differed for subgroups, with the caveat that tests for heterogeneous treatment effects across contexts are usually less statistically powerful than those for main effects; this adds to the problem of affirming the null hypotheses to caution the interpretation of the lack of significant interactions between the intervention effect and contexts.

Amount of Professional Development

Although the main conclusions concern a curriculum, a second implication is that substantial professional development may be necessary to achieve a high-quality implementation of that curriculum. Merely adopting a new curriculum is one of the most common, but usually unsuccessful, external interventions (Ball & Cohen, 1999). We incorporated professional development. The 34 hours of focused group work and approximately 16 hours of in-class coaching of the present study is substantially more than offered to most teachers. For example, only 6% of elementary school teachers participate in mathematics professional development for more than 24 hours over a year (Birman et al., 2007). The total of 50 hours in this study is consistent with previous research (cf. an average of 53 hours of teaching training yielding an average effect size of .53 on student math achievement in Yoon, Duncan, Lee, Scarloss, & Shapley, 2007), suggesting a minimal duration for effective professional development.

Usefulness of the Curriculum Research Framework

The results also provide additional (Clements & Sarama, 2007c) "proof-of-concept" support for our theoretical framework for research-based curricula (CRF), including three categories, A Priori Foundations, Learning Model, and Evaluation, and 10 phases described previously, which extends and particularizes theories of curriculum research (Clements, 2002, 2007). An implication is that such synthesis of curriculum development as a scientific enterprise and mathematics education research may help reduce the separation of research and practice and produce results that are applicable not only to researchers but to practitioners (parents, teachers, and teacher educators), administrators, policy makers, and curriculum and software developers. A caveat is that this study involved a moderate number of mostly volunteer teachers located in proximity to the researchers.

Efficacy of the Learning Trajectories Construct

The larger effects of the *Building Blocks* curriculum than the comparison curriculum support the distinct contribution of research-based learning trajectories. Given that the content coverage of the two were closely matched, significant differences favoring the *Building Blocks* curriculum may be the result of the instructional strategies and the pedagogical content knowledge embedded in its learning trajectories. These results intimate that others will find learning trajectories a useful construct in future research, curriculum development, and professional development efforts. Learning trajectories have several advantages, including that (a) the developmental progressions provide benchmarks for formative assessments, especially useful for children who are low performers; (b) these progressions can form a foundation for future curriculum development; (c) these progressions can be tested and refined in future research projects to serve goals in cognitive psychology and mathematics education; (d) the other main component of the learning trajectories, the sequence of instructional tasks, can be adopted for teaching and professional development efforts; and (e) subsets of these instructional sequences can be tested against others to increase the efficacy of the learning trajectory. *Building Blocks* interweaves different learning trajectories; therefore, these two characteristics, use of learning trajectories and the particular instantiation of them as interwoven throughout the curriculum, are intentionally confounded, and thus their effects can not be distinguished. Interweaving the learning trajectories is supported by theory and research, as discussed previously. Nevertheless, although inconsistent with the CRF model, it would be possible to package learning trajectories, or topics, into conventional, sequential units, permitting future research to ascertain the degree to which positive effects can be attributed to learning trajectories without interweaving, to interweaving topics without using developmental progressions, or to the synthesis of the two, as instantiated in the *Building* Blocks curriculum. Teachers in the *Building Blocks* curriculum also used a greater number of activities to address developmental levels, but the data from this study can not separate the impact of the number of activities per se from their use within learning trajectories.

We are presently studying whether these effects can be maintained with large numbers of schools, teachers, and students and with great geographical and contextual diversity and whether the effects are sustainable over several years. Also, other research teams should test the CRF and its core of learning trajectories, including comparing it to alternatives, to evaluate the effectiveness of the model per se.

Note

This paper was based upon work supported in part by the National Science Foundation (NSF) under Grant No. ESI-9730804 to D. H. Clements and J. Sarama, "Building Blocks—Foundations for Mathematical Thinking, Pre-Kindergarten to Grade 2: Research-Based Materials Development," and in small part by the Institute of Educational Sciences (IES; U.S. Department of Education, under the Interagency Educational Research Initiative, or IERI, a collaboration of the IES, NSF, and National Institute of Child Health and Human Development) under Grant No. R305K05157 to D. H. Clements, J. Sarama, and J. Lee, "Scaling Up TRIAD: Teaching Early Mathematics for Understanding with Trajectories and Technologies." Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies. The curriculum evaluated in this research has since been published by the authors, who thus have a vested interest in the results. An external auditor oversaw the research design, data collection, and analysis, and five researchers independently confirmed findings and procedures. The authors, listed alphabetically, contributed equally to the research.

References

Arnold, D. H., Fisher, P. H., Doctoroff, G. L., & Dobbs, J. (2002). Accelerating math development in Head Start classrooms: Outcomes and gender differences. *Journal of Educational Psychology*, *94*, 762–770.

Balfanz, R. (1999). Why do we teach young children so little mathematics? Some historical considerations. In J. V. Copley (Ed.), *Mathematics in the early years* (pp. 3–10). Reston, VA: National Council of Teachers of Mathematics.

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

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

Birman, B. F., LeFloch, K. C., Klekotka, A., Ludwig, M., Taylor, J., Walters, K., et al. (2007). *State and local implementation of the No Child Left Behind Act, volume II—Teacher quality under NCLB: Interim report.* Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

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

Bodovski, K. (2006, April). *Instruction, student engagement, and mathematics learning in elementary school.* Paper presented at the American Educational Research Association, San Francisco.

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

Brosterman, N. (1997). *Inventing kindergarten*. New York: Harry N. Abrams.

Bryant, D. M., Burchinal, M. R., Lau, L. B., & Sparling, J. J. (1994). Family and classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly*, *9*, 289–309.

Campbell, P. F., & Silver, E. A. (1999). *Teaching and learning mathematics in poor communities.* Reston, VA: National Council of Teachers of Mathematics.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380.

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

Clements, D. H. (1984). Training effects on the development and generalization of Piagetian logical operations and knowledge of number. *Journal of Educational Psychology*, *76*, 766–776.

Clements, D. H. (2002). Linking research and curriculum development. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 599–636). Mahwah, NJ: Lawrence Erlbaum.

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., & Conference Working Group. (2004). Part one: Major themes and recommendations. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 1–72). Mahwah, NJ: Lawrence Erlbaum.

Clements, D. H., & Sarama, J. (2003). *DLM Early Childhood Express math resource guide.* Columbus, OH: SRA/McGraw-Hill.

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). Hypothetical learning trajectories [Special issue]. *Mathematical Thinking and Learning*, *6*(2).

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

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). Early childhood mathematics learning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 461–555). New York: Information Age.

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

Clements, D. H., Wilson, D. C., & Sarama, J. (2004). Young children's composition of geometric figures: A learning trajectory. *Mathematical Thinking and Learning*, *6*, 163–184.

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

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis*, *24*, 175–199.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., et al. (Eds.). (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements.* San Francisco: Jossey-Bass.

Denton, K., & West, J. (2002). Children's reading and mathematics achievement in kindergarten and first grade. Retrieved December 1, 2002, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002125

Farran, D. C., Silveri, B., & Culp, A. (1991). Public preschools and the disadvantaged. In L. Rescorla, M. C. Hyson, & K. Hirsh-Pase (Eds.), *Academic instruction in early childhood: Challenge or pressure? New directions for child development* (pp. 65–73). San Francisco: Jossey-Bass.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher*, *31*(8), 4–14.

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.

Fuson, K. C. (1988). *Children's counting and concepts of number.* New York: Springer-Verlag.

Fuson, K. C., Smith, S. T., & Lo Cicero, A. (1997). Supporting Latino first graders' ten-structured thinking in urban classrooms. *Journal for Research in Mathematics Education*, *28*, 738–760.

Galván Carlan, V. (2000). *Development of an instrument to assess the use of developmentally appropriate practices in teaching mathematics in early childhood classrooms.* Unpublished doctoral dissertation, University of Houston.

Galván Carlan, V., & Copley, J. V. (2000). *Early childhood mathematics profile observational checklist.* Houston, TX: University of Houston.

Griffin, S. (2004). Number Worlds: A research-based mathematics program for young children. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 325–342). Mahwah, NJ: Lawrence Erlbaum.

Griffin, S., & Case, R. (1997). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education*, *3*(1), 1–49.

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

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

Holloway, S. D., Rambaud, M. F., Fuller, B., & Eggers-Pierola, C. (1995). What is "appropriate practice" at home and in child care? Low-income mothers' views on preparing their children for school. *Early Childhood Research Quarterly*, *10*, 451–473.

Hord, S., Rutherford, W., Huling-Austin, L., & Hall, G. (1987). *Taking charge of change.* Alexandria, VA: Association for Supervision and Curriculum Development.

Horizon Research. (2001). *Classroom observation protocol.* Chapel Hill, NC: Author.

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

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

Klein, A., Starkey, P., & Ramirez, A. B. (2002). *Pre-K mathematics curriculum.* Glenview, IL: Scott Foresman.

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

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate.* Washington, DC: Economic Policy Institute.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., et al. (2000). *TIMSS 1999 international mathematics report*. Boston: International Study Center, Boston College, Lynch School of Education.

National Association for the Education of Young Children. (1991). *Early childhood classroom observation—National Academy of Early Childhood Programs* (Rev. ed.). Washington, DC: Author.

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. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: Mathematical Sciences Education Board, Center for Education, Division of Behavioral and Social Sciences and Education, National Academies Press.

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

Piaget, J., & Szeminska, A. (1952). *The child's conception of number*. London: Routledge/Kegan Paul.

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, *64*(6), 916–924.

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist*, *53*, 109–120.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, *2*, 173–185.

Razel, M., & Eylon, B.-S. (1991, July). *Developing mathematics readiness in young children with the Agam Program*. Paper presented at the 15th conference of the International Group for the Psychology of Mathematics Education, Genoa, Italy.

Reeves, D. B. (2002, May 8). Galileo's dilemma. *Education Week*, p. 44.

Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education*, *32*, 368–398.

Rohrer, D., & Taylor, K. (2006). The effect of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*, 1–16.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Sarama, J. (2004). Technology in early childhood mathematics: *Building Blocks™* as an innovative technology-based curriculum. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 361–375). Mahwah, NJ: Lawrence Erlbaum.

Sarama, J., & Clements, D. H. (2002). *Building Blocks* for young children's mathematical development. *Journal of Educational Computing Research*, *27*(1/2), 93–110.

Sarama, J., & Clements, D. H. (in press). *Research-based Elementary Math Assessment*. Columbus, OH: SRA/McGraw-Hill.

Sarama, J., Clements, D. H., & Vukelic, E. B. (1996). The role of a computer manipulative in fostering specific psychological/mathematical processes. 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. 567–572). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Saxe, G. B., Guberman, S. R., & Gearhart, M. (1987). Social processes in early number development. *Monographs of the Society for Research in Child Development*, *52*(2, Serial 216).

Secada, W. G. (1992). Race, ethnicity, social class, language, and achievement in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 623–660). New York: Macmillan.

Senk, S. L., & Thompson, D. R. (2003). *Standards-based school mathematics curricula. What are they? What do students learn?* Mahwah, NJ: Lawrence Erlbaum.

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development.* Washington, DC: National Academy Press.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking.* New York: Oxford University Press.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*(2), 114–145.

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

Starkey, P., Klein, A., Chang, I., Qi, D., Lijuan, P., & Yang, Z. (1999). Environmental supports for young children's mathematical development in China and the United States. Albuquerque, NM: Society for Research in Child Development.

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.

Teaching Strategies. (2001). *A checklist for assessing your program's implementation of the Creative Curriculum for early childhood* (3rd ed.). Washington, DC: Author.

Wright, R. J., Stanger, G., Stafford, A. K., & Martland, J. (2006). *Teaching number in the classroom with 4-8 year olds.* London: Paul Chapman/Sage.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues and Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Manuscript received March 28, 2007
Manuscript accepted November 19, 2007