Journal for Research in Mathematics Education
2011, Vol. 42, No. 2, 127–166

Mathematics Learned by Young Children in an Intervention Based on Learning Trajectories: A Large-Scale Cluster Randomized Trial

Douglas H. Clements, Julie Sarama, Mary Elaine Spitler,
Alissa A. Lange, and Christopher B. Wolfe
University at Buffalo, State University of New York

This study employed a cluster randomized trial design to evaluate the effectiveness of a research-based intervention for improving the mathematics education of very young children. This intervention includes the *Building Blocks* mathematics curriculum, which is structured in research-based learning trajectories, and congruous professional development emphasizing teaching for understanding via learning trajectories and technology. A total of 42 schools serving low-resource communities were randomly selected and randomly assigned to 3 treatment groups using a randomized block design involving 1,375 preschoolers in 106 classrooms. Teachers implemented the intervention with adequate fidelity. Pre- to posttest scores revealed that the children in the Building Blocks group learned more mathematics than the children in the control group (effect size, $g = 0.72$). Specific components of a measure of the quantity and quality of classroom mathematics environments and teaching partially mediated the treatment effect.

Key words: Curriculum; Early childhood; Early number learning; Equity/diversity; Geometry; Instructional technology; Preschool/primary; Teaching effectiveness

Recent national reports highlight the importance of effective mathematics education in preschool (National Mathematics Advisory Panel, 2008; National Research Council, 2009). We designed the *Building Blocks* preschool mathematics curriculum (Clements & Sarama, 2007a) as a set of tools that would enable all young children to build a solid foundation for mathematics, and especially that would increase the mathematical knowledge of children from low-resource communities. This study evaluates the effectiveness of these tools when introduced to multiple

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305K05157 to the University at Buffalo, State University of New York, "Scaling Up TRIAD: Teaching Early Mathematics for Understanding with Trajectories and Technologies." The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education. It should be noted that, although the research is concerned with a complete research-based intervention, not just a particular curriculum, the curriculum evaluated in this research has been published by the authors, who thus could have a vested interest in the results. However, an external auditor oversaw the research design, data collection, and analysis, and five researchers independently confirmed findings and procedures.

urban school districts on a large scale, analyzing the specific mathematics concepts and skills that young children learned in these classes.

The rationale for this close examination of an instantiation of a research-based curriculum derives from the confluence of research and educational needs in mathematics and early childhood. Although all citizens need broad mathematical understandings, U.S. students' mathematical proficiency has been evaluated as "low," especially compared to that of students of other countries (Campbell & Silver, 1999; Mullis et al., 2000; National Mathematics Advisory Panel, 2008; National Research Council, 2001a). Moreover, children from low-resource communities who are members of linguistic and ethnic minority groups demonstrate significantly lower levels of achievement than children from higher-resource, nonminority communities (Denton & West, 2002; National Research Council, 2001b; National Research Council, 2009; Natriello, McDill, & Pallas, 1990; Sarama & Clements, 2009). High-quality experiences in preschool result in greater competence in a variety of domains upon entry into kindergarten (Christian, Morrison, Frazier, & Massetti, 2000; Clements & Sarama, 2008a; Magnuson, Meyers, Ruhm, & Waldfogel, 2004; National Research Council, 2001b; National Research Council and Institute of Medicine, 2000; Sarama & Clements, 2009). Further, the same high-quality experiences may benefit low-income children more because they have fewer educational opportunities in their homes (Carneiro & Heckman, 2003; Raudenbush, 2009); such effects have been reported as strongest for children from low-resource communities and for children whose parents have less formal education, with benefits persisting into high school (Brooks-Gunn, 2003). Gains from early education programs, however, are not guaranteed. A large-scale evaluation of Head Start found limited effects on early mathematics skills for 3-year-olds, but no effects for 4-year-olds and no evidence of impact on mathematics at the end of kindergarten or grade 1 (U.S. Department of Health and Human Services, Administration for Children and Families, 2010). Such disparate findings suggest that more research is needed to clarify the relationships between various preschool experiences and later academic achievement.

Further, there is a need to evaluate the effects of curriculum-based interventions. Curricula rarely are evaluated scientifically (Clements, 2007); for example, fewer than 2% of research studies in mathematics education have concerned the effects of written curricula (Senk & Thompson, 2003). This is particularly unfortunate because the effect sizes for curricular interventions, according to a policy statement by Whitehurst (2009), generally may be larger than for popular reforms such as charter schools, reconstituting the teacher workforce, preschool, and common standards. From a human capital perspective, the greatest benefit is gained from interventions with the youngest children (Carneiro & Heckman, 2003; Heckman, 2003; Krueger, 2003). Whitehurst (2009) argues that only expensive early childhood programs have been proven effective and that "scarce resources need to be allocated to get the biggest bang for the buck" (p. 6). However, he does not address the integration of curriculum and early childhood interventions. Collectively, these analyses and policy statements suggest that theoretically grounded and empirically grounded curriculum-based interventions in early childhood may constitute an efficacious and cost-effective route to raising achievement in low-resource communities.

Such a hypothesis raises the issue of a definition of "research-based" curricula. The *Building Blocks* curriculum was developed and evaluated using the comprehensive Curriculum Research Framework (CRF) (Clements, 2007). Descriptions of the implementation of the CRF in the case of *Building Blocks* are included in two previous reports, an initial summative research study yielding effect sizes (Cohen's *d*) between 1 and 2 (Clements & Sarama, 2007c) and a final-phase, proof-of-concept scale-up study (Clements & Sarama, 2008a), in which the scores of the Building Blocks group increased significantly more than the scores of a comparison group that received another preschool mathematics curriculum (effect size, 0.47) and more than a control group (effect size, 1.07). The research reported here is a full-scale study of scale-up of this *Building Blocks* curriculum.

At the core of the CRF are empirically grounded learning trajectories (cf. Simon, 1995). We define learning trajectories as "descriptions of children's thinking and learning in a specific mathematical domain and a related, conjectured route through a set of instructional tasks designed to engender those mental processes or actions hypothesized to move children through a developmental progression of levels of thinking, created with the intent of supporting children's achievement of specific goals in that mathematical domain" (Clements & Sarama, 2004b, p. 83). Building on the work of such projects as Cognitively Guided Instruction (CGI) (Carpenter, Franke, Jacobs, Fennema, & Empson, 1998) and thousands of research studies (see Clements & Sarama, 2009; Sarama & Clements, 2009), *Building Blocks* learning trajectories are not simply "educated guesses" but are based on empirically supported developmental progressions (more so for more heavily researched topics, of course). 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 both with and without the computer, the structures of which correspond to the levels of this developmental progression (Clements & Sarama, 2007c; Sarama et al., 1996).

Thus, the teacher has a well-formed and specific set of expectations about children's ways of learning and a likely pace along a path that includes central, worthwhile ideas. We believe that learning trajectories are an effective way to both motivate and support the use of the empirically supported instructional practice of formative assessment (National Research Council, 2009; see also National Mathematics Advisory Panel, 2008, for reference to research on formative assessment).

We agree that these are hypothetical learning trajectories (Simon, 1995). That is, the instantiation by the teacher and then by the interaction of the teacher, students, and activities is a determinant of the nature, quality, and effectiveness of the curriculum's learning trajectories. To address criticisms of and research on the development and professional use of curriculum materials (e.g., Ball & Cohen, 1996; Davis & Krajcik, 2005; Sarama & DiBiase, 2004), the *Building Blocks* curriculum focuses not only on children's development of mathematics but also on teachers' professional development and curriculum enactment. With a common curriculum and knowledge of mathematical learning trajectories, we intended to provide teachers and children with a research-based, shared, systematic practice that is more effective and more amenable to scientifically based improvement than private, idiosyncratic practice (Raudenbush, 2009).

As stated, previous studies have supported the efficacy of the *Building Blocks* curriculum and this approach to its implementation (Clements & Sarama, 2007c, 2008a). However, each of these evaluations involved a relatively small number of teachers (all of whom were volunteers) and students. The purpose of this study was to evaluate whether a similar implementation of the curriculum would have comparable effects at a large scale and to document in detail what mathematics is learned in the contexts of the Building Blocks and control classrooms. The CRF posits that a valid scientific curriculum development program should address two basic issues, effect and conditions, in three domains: practice, policy, and theory (see Clements, 2007, Table 1, p. 39). Therefore, we examine not only the intervention's effect on specific learning goals but also whether any such effects are mediated by particular pedagogical practices and whether they are equivalent for various subpopulations, including subpopulations defined by different school contexts (socioeconomic status [SES] and percent limited English proficiency [LEP]) or child characteristics (e.g., gender, having an Individualized Education Plan, or IEP).

METHOD

We used a multisite cluster randomized trial (CRT) experimental design that enabled a test of the generalizability of the *Building Blocks* intervention's impact over the varied settings in which it was implemented. We employed hierarchical linear models (*HLM*s) to measure the effects of the intervention on students' mathematics performance and to account for possible variations of the effects among varied contexts.

Participants and Contexts

We initially contacted 10 urban public school districts; two met the criteria of (a) serving ethnically diverse populations who live mainly in poverty; (b) having a large number of prekindergarten (pre-K) classrooms within elementary schools, with self-contained feeder patterns (a history of preschoolers continuing their education in that school); (c) willingness to ensure that each pre-K classroom would have two Internet-enabled computers; and (d) willingness to have schools randomly assigned to treatments (thus, not having a single mandated pre-K mathematics curriculum). All schools in which pre-K teachers had not previously been involved in *Building Blocks* research or development projects were included. Both superintendents decided to adopt the *Building Blocks* curriculum, so 42 schools meeting

Table 1
Demographics of Children and Schools in Study

	Experimental	Control
Number of children (*n*)	927	378
Race/Ethnicity		
American Indian	17 (2%)	7 (2%)
African American	513 (55%)	182 (48%)
Asian/Pacific Islander	29 (3%)	19 (5%)
Hispanic	182 (20%)	100 (27%)
White, non-Hispanic	181 (20%)	67 (18%)
Other	5 (<1%)	3 (1%)
Gender		
Male	452 (49%)	188 (50%)
Female	475 (51%)	190 (50%)
IEP status		
Yes	80 (9%)	25 (7%)
No	847 (91%)	353 (93%)
Number of schools	26	16
Average percentage free/reduced lunch	85%	85%
Average percentage LEP	13%	15%
Average number of study (pre-K) teachers within schools	1.71	2.13
Number of study (pre-K) classes	72	34
Average number of study (assessed) children per class	12.9	11.2

Notes. AA = African American, A/P = Asian/Pacific Islander, NA = Native American, H = Hispanic, W = White non-Hispanic (records allowed no further categorical breakdowns). LEP = percent Limited English Proficiency in school; SES = percent free/reduced lunch in school. Percentages are rounded to whole numbers for readability.

those criteria were involved as part of a “phased adoption” (i.e., principals and teachers were not volunteers, and all control schools received curricula after the end of data collection). Schools within each district were ordered on the basis of their average scores on state-based mathematics achievement tests. Using a randomized block procedure and a table of random numbers, we then publicly (supervised by two school administrators and three staff members) assigned each school to one of three treatment groups. The two treatment groups differed only in that one included a follow-through component that was to be implemented when the children moved to kindergarten and first grade; therefore, at the pre-K level, the focus of this study, there were only two distinct groups, with 2/3 of the schools in the Building Blocks intervention group and 1/3 in the control group. Statistical power calculations suggested the inclusion of 12 children from each pre-K classroom. To compensate for attrition, we set a maximum of 15. From the pool of all kindergarten-intending children (in the entry range for kindergarten—many classrooms had both 3-year-olds and 4-year-olds) who returned consent forms, we randomly selected up to 15 children (attrition resulted in the number of children who completed all assessments ranging from 5 to 15 per class).

Table 1 shows the diverse populations at the school level, using data from school records as well as parent and teacher questionnaires. Ninety-nine percent of the teachers were female. About 89% had at least a master's degree. They had an average of 17.0 years of teaching experience (Building Blocks, 16.9; control, 17.2), and their average number of years with prekindergarten teaching experience was 6.8 (Building Blocks, 6.6; control, 7.1). About 94% believed themselves to have the support of their principals. As a measure of morale, 79% of the teachers believed that "staff members in this school generally have school spirit" (Building Blocks, 76%; control, 85%).

Curricula for Children and Teachers

Building Blocks (Clements & Sarama, 2007a) is a National Science Foundation–funded[1] mathematics curriculum designed using a comprehensive Curriculum Research Framework (Clements, 2007) to address numeric/quantitative and geometric/spatial ideas and skills. Woven throughout are mathematical subthemes, such as sorting and sequencing, as well as mathematical processes. General processes include communicating, reasoning, representing, and problem solving and the overarching mathematizing. Specific mathematical processes include number and shape composition and patterning. These were determined to be critical mathematical building blocks (the same body of research and expertise guided the consonant *Curriculum Focal Points*, National Council of Teachers of Mathematics, 2006).

Building Blocks' instructional approach "is to find the mathematics in, and develop mathematics from, children's experiences and interests" (Clements & Sarama, 2007a). Children are guided to extend and mathematize their everyday activities, from block building to art to songs to puzzles, through sequenced, explicit activities (whole group, small group, centers, including a computer center, and "throughout the day"). Thus, off-computer and on-computer activities are designed based on children's experiences and interests, with an emphasis on supporting the development of mathematical activity. More detailed descriptions of *Building Blocks* appear in Clements and Sarama (2004a, 2007a, 2007c) (in addition, see http://UBBuildingBlocks.org).

The *Building Blocks* learning trajectories were designed to develop teachers'

[1]Based in part upon work supported by the National Science Foundation Research Grant ESI-9730804, "Building Blocks—Foundations for Mathematical Thinking, Pre-Kindergarten to Grade 2: Research-Based Materials Development."

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

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

In this project, teachers participated in 8 days of professional development during the school day in the 1st year (a “no stress, gentle introduction” to the curriculum with no data collection by researchers) and 5 days in the 2nd year focused on the learning trajectories for each mathematical topic. Training addressed each of the three components of the learning trajectories. To understand the goals, teachers learned core mathematics concepts and procedures for each topic. For example, they studied the system of verbal counting based on cycling through 10 digits and the concept of place value (based on content similar to that presented in National Research Council, 2009). To understand the developmental progressions of levels of thinking, teachers studied multiple video segments illustrating each level and discussed the mental “actions on objects” that constitute the defining cognitive components of each level. To understand the instructional tasks, teachers studied the tasks, and they viewed, analyzed, and discussed video of the enactments of these tasks in classrooms.

Each of these professional development contexts used the software application Building Blocks Learning Trajectories (BBLT), which presented and connected all components of the innovation. BBLT provided scalable access to the learning trajectories via descriptions, videos, and commentaries. The two sequential aspects of the learning trajectories, the developmental progressions of children's thinking and connected instructional tasks, are linked to each other (see Clements & Sarama, 2008a, p. 463, or see http://UBTRIAD.org). Discussions of BBLT classroom videos made explicit how such practice exemplified research-based principles.

The professional development sessions were sequenced following the *Building Blocks* curriculum. Throughout this study, teachers learned how to use the learning trajectories as a basis for formative assessment, a key to high-quality teaching (see National Mathematics Advisory Panel, 2008). Formative assessment has been shown to be particularly difficult for teachers to enact without substantial support (Foorman et al., 2007). In the professional development sessions, teachers discussed and practiced how to interpret children's thinking and select appropriate instructional tasks for the class (e.g., compacting the curriculum if most children can learn it at a faster pace) and for individuals (e.g., assigning children to small groups or modifying activities within groups to match instructional tasks to developmental levels of individual children).

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

Curricula in control classrooms. Information on curricula used in control classrooms was gathered via (a) administrator surveys with all principals conducted through telephone interviews (all but one was interviewed), (b) written teacher questionnaires (completed by all but three teachers), (c) telephone interviews with about one fourth of the teachers using a protocol designed to obtain information on curriculum practices, (d) informal visits with all teachers (except one), and (e) two formal observations of each teacher's mathematics activities (see the following sections on measures).

In both districts, there was a greater focus on mathematics during the study period than there had been in prior years, due to the introduction of new comprehensive programs that included mathematics, during Year 1 of the project. The first district implemented *Where Bright Futures Begin* (Bredekamp, Morrow, & Pikulski, 2006). The curriculum features 10 thematic segments (e.g., *Animals Everywhere*), each consisting of 3 weeks of theme-related instruction. The mathematics component included nine topics: geometry and spatial sense, patterns, time concepts, measurement, classification and data collection, numbers and operations, problem solving, reasoning, and communication. Examples of number learning goals were "counts and builds sets of one to five objects" and "uses one-to-one correspondence to arrange and compare sets." Examples of measurement goals included "compares size, length, capacity, weight in natural situations" and "measures length of objects." Mathematics materials included 34 mathematics concept cards, as well as everyday mathematical classroom objects, such as counters and cubes. Mathematics activities were taught primarily during small group time, although sometimes during whole group instruction. Length Treasure Hunt provides an example. Children take classroom objects from a bag, compare lengths while working with partners, and then identify which objects are about the same length and which are not, using the terms *shorter than* and *longer than*. Suggestions for differentiated instruction ("Challenge" or "Extra Support") were offered, as were highlighted connections between mathematics and literacy, with an emphasis on mathematics vocabulary. The teacher manual also offered teacher support for instructing English Language Learners and for formative assessment of children (see Houghton Mifflin Harcourt, n.d.). Professional development for *Where Bright Futures Begin* was provided for teachers three times during the project's second year, each time with an emphasis on literacy. Interviews and classroom observations of local-site control teachers revealed instruction on mathematics topics taught in the *Building Blocks* curriculum, including counting, recognizing number, comparing number, shape, comparing shape, representing shapes, measurement, patterning and sorting, classifying, and graphing. Teachers reported minimal teaching on composition of number, composition of shape, adding and subtracting, transformations, and probability.

The second district had begun, on a staggered basis, the implementation of a new comprehensive curriculum, *Opening the World of Learning* (*OWL*) (Schickedanz & Dickinson, 2005), which was designed for full-day implementation, with components added to language and literacy, including mathematics, science, social studies, art, and physical, social, and emotional development. Some distal-site control teachers used this curriculum. *OWL* mathematics activities were presented as small group activities, such as Watch Me Count (children watch and then duplicate a teachers' counting of a small number of objects, then of sets with larger numbers of objects). Components included suggested vocabulary, with procedures provided for extra support as well as extension activities. Topics included number concepts, number words, one-to-one correspondence, cardinality, basic computation, geometry, and measurement; domains consisted of number sense, numeration, spatial sense, measurement, geometry, and patterns (Schickedanz & Dickinson, 2005). Approximately six professional development sessions on *OWL* were provided between fall 2005 and spring 2007. Five of the 10 control teachers (not using *OWL*) reported adapting mathematics activities from the kindergarten curriculum *Investigations in Number, Data, and Space* (Economopoulos, Murray, Eston, & Kliman, 2008); one mentioned combining the kindergarten-level *Investigations* with some mathematics from *OWL;* three mentioned combining *Investigations* with *DLM Early Childhood Express* (Schiller, Clements, Sarama, & Lara-Alecio, 2003), whose mathematics component was an earlier version of the *Building Blocks* curriculum. (Project leaders were aware of this contamination threat, but not the extent of *DLM* infusion, which is addressed in a subsequent section.) Control teachers reported teaching counting, number recognition, comparing and composing number, and addition and subtraction. They also taught shape identification, shape comparison, shape composition, measurement, patterning, and sorting. They reported teaching the following topics with less emphasis: transformations, probability, and reasoning/problem solving.

Measures

Children's mathematical knowledge. The Research-based Elementary Math Assessment (REMA) (see also Clements, Sarama, & Liu, 2008; Sarama & Clements, 2011) measures core mathematical abilities of preschool children using an individual interview format, with standardized protocol and scoring procedures. Abilities are assessed according to theoretically based and empirically based developmental progressions that underlie the Building Blocks project's learning trajectories. Developmental progressions in number include verbal counting, object counting, subitizing, number comparison, number sequencing, connection of numerals to quantities, number composition and decomposition, adding and subtracting, and place value. These domains help distinguish between children who have not constructed true number concepts and those who have. Geometry progressions include shape recognition, shape composition and decomposition, congruence, construction of shapes, and spatial imagery, as well as geometric measurement and patterning. The REMA defines mathematical competence as a latent trait in item response theory (IRT), yielding a score that locates children on a common ability scale with a consistent, justifiable metric (Wright & Stone, 1979). All items are ordered by Rasch item difficulty; children stop the assessment after four consecutive errors. Explicit protocols and procedures exist for administration, videotaping, coding, and scoring and for staff training on all aspects. Each item is coded by two trained coders for accuracy and, when relevant, for solution strategy. Any discrepancies were resolved via consultation with the senior researchers. Continuous coder calibration militated against drift. Previous analysis of the assessment data showed that its reliability ranged from 0.75 to 0.94 on the subtests and 0.93 to 0.94 on the total test scores (see Clements et al., 2008, for full details on content and concurrent validity); on the present population, the reliability was 0.92. For the present study, inferential statistics were performed on Rasch scores computed for the total instrument. The sum of raw scores was computed for items within each mathematical topic for descriptive purposes.

Teachers' classroom practices (*e.g., implementation fidelity*), *knowledge, and beliefs*. Three instruments provided data on teachers' knowledge, beliefs, and practice, an essential measure given the need to consider instruction as the proximal cause of student learning (Raudenbush, 2008). The teacher questionnaire measured teacher's self-reported knowledge, beliefs, and practices pertaining to early childhood mathematics, including sections on demographics, education and experience, mathematics goals, children's learning, and teaching practices.

Two related observational instruments described and measured teachers' classroom practices. Our observational instruments were designed to assess the "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 about the characteristics and teaching strategies of effective teachers of early childhood mathematics (e.g., Clarke & Clarke, 2004; Clements & Sarama, 2007b; Fraivillig, Murphy, & Fuson, 1999; Galván Carlan, 2000; Horizon Research, Inc., 2001; Teaching Strategies, Inc., 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 titled 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 Organization, Teaching Approaches, and 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 it and COEMET were designed to document more deeply how mathematics is taught and what happens in each classroom. Evidence of their validity can be found in earlier studies (Clements & Sarama, 2008a; Sarama, Clements, Starkey, Klein, & Wakeley, 2008); for example, the COEMET correlated significantly with child gain scores ($r = .50$) (Clements & Sarama, 2008a).

The Fidelity instrument contains 52 items and responses to most are on 5-point Likert scales. An example of an item unique to the Fidelity measure in the Organization, Teaching Approaches, and 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)." Thus, the Fidelity instrument includes sections for each component of the implemented curriculum, such as a specific small-group activity. Only activities prescribed in the implemented curriculum are evaluated, and ratings are made in reference to the printed curriculum. To observe an activity from each component of each curriculum, visits were approximately 1 hour in duration. Interrater reliability, computed via simultaneous classroom visits by pairs of observers (10% of all observations, with pair memberships rotated) averaged 91% in previous research (Clements & Sarama, 2008a) and 95% in the present study. High reliability was reported in previous research ($\alpha = .90$; Clements & Sarama, 2008a).

The COEMET measures the quality of the mathematics environment and activities with a half-day observation and can be used with different curricula. Thus, it allows for experimental–control group contrasts. There are 28 items, all but 4 of which have similar 5-point Likert scales. An example of one of the three items in a section unique to this measure, Classroom Culture (subsection: Personal Attributes of the Teacher), is "The teacher appeared to be knowledgeable and confident about mathematics (i.e., demonstrated accurate knowledge of mathematical ideas and procedures, demonstrated knowledge of connections between, or sequences of, mathematical ideas)." Assessors spend no less than 1 half-day in the classroom, for example, from before the children arrived until the end of the half-day or until lunch (about half were conducted after lunch to the end of the school day, if that was the period during which mathematics was taught). All mathematics activities are observed and evaluated, without reference to any printed curriculum (i.e., assessors are not told what curriculum is present, are not familiar with the mathematics curricula involved, and thus are intended to be naïve as to treatment group). As shown in Table 2, the COEMET has three main sections: Classroom Culture, Specific Math Activity(ies) (SMAs), and Other Classroom Elements (e.g., number of computers on which children are working). We also compute a total score, consisting of the sum of the Likert items from the Classroom Culture and SMAs sections. Reliability in the present study was computed for the total instrument ($\alpha = .95$) as well as for Classroom Culture ($\alpha = .74$) and SMAs ($\alpha = .90$). Assessors complete the Classroom Culture and some classroom elements sections once to reflect their entire observation. They complete an SMA form for each observed mathematics activity, defined as one conducted intentionally by the teacher involving several interactions with one or more children, or conducted intentionally to develop mathematics knowledge (this would not include, for instance, a single, informal statement about number or shape).

Interrater reliability for the COEMET, computed via simultaneous classroom visits by pairs of observers (10% of all observations, with pair memberships rotated) averaged 0.88 in previous research (Clements & Sarama, 2008a) and 0.80 in the present study; 99% of the disagreements were the same polarity (i.e., if one was agree, the other was strongly agree) in both. Cronbach's alpha (interitem correlations) for the two instruments ranged from 0.95 to 0.97 in previous research (Clements & Sarama, 2008a; Sarama, Clements, Starkey, Klein, & Wakeley, 2008) and 0.95 in the present study.

Procedures

The 1st year was a pilot/training year, because our previous experience (Clements & Sarama, 2008a; Sarama et al., 2008) and others' research suggested that teachers often need at least 1 year of experience before completely and effectively implementing a curriculum (Berends, Kirby, Naftel, & McKelvey, 2001; Cobb, McClain, de Silva, & Dean, 2003; Weiss, 2002). Assessors who would conduct child assessments were recruited and trained for 4 days during Year 1, including reading the REMA, practicing administration, and sending videos of the latter sessions to the University at Buffalo for evaluation and feedback (only those trainees meeting the

Table 2
Means and Standard Deviations of Classroom Observation Measures by Treatment Group

	Min	Max	Experimental		Control		Total	
			n	*M* (*SD*)	*n*	*M* (*SD*)	*N*	*M* (*SD*)
Fidelity of Implementation								
Total score	–22	64.50	67	37.4 (15.16)				
COEMET								
Classroom Culture subscore	18.5	41.5	72	36 (3.9)	34	31 (4.7)	106	34 (4.7)
SMA subscore	46.0	87.9	72	72 (4.7)	34	68 (6.8)	106	71 (5.8)
Other Classroom Elements								
Total number of SMAs (*M*)	1.5	14	72	7 (2.3)	34	5 (1.7)	106	6 (2.3)
Time on task	8.2	92.5	72	33 (15.5)	34	27 (12.3)	106	31
(min/day)								(14.7)
Number of computers working for children	0.0	6.0	72	3 (1.2)	34	1 (1.3)	106	2 (1.3)

criteria of two consecutive errorless administrations were certified). REMA assessors were also naïve to children's (i.e., the school's) assignment to treatment.

All mentors and teachers from Building Blocks schools participated in professional development after teachers completed the teacher questionnaire, which were the only data collected in Year 1. All mentors and teachers from Building Blocks schools participated in the course. Teachers were encouraged to engage their children in all curriculum components immediately after the first professional development session. In the context of the course, teachers continued to learn about learning trajectories, using the BBLT (Figure 1) and to engage in discussions of these and correlated text from the *Building Blocks* curriculum.

In Year 2, complete data collection was implemented. We administered the REMA as a pretest and posttest to all eligible pre-K children. Mentors collected fidelity data in all Building Blocks classrooms twice to assess the degree to which teachers implemented the letter and spirit of the *Building Blocks* curriculum. COEMET observers were trained in late summer. Procedures were similar to those for the Fidelity instrument, but observers were naïve to the school's assignment to treatment. COEMET data were collected once in late fall and once in early spring in all classrooms.

Teachers were asked to engage their children in all curriculum components after pretesting was completed for their classes. Throughout that year's professional development sessions, teachers continued to study the learning trajectories, including discussions of how they conducted various curricular activities the previous year. As part of this work, teachers brought case studies of particular situations that occurred in their classrooms to the group to facilitate these discussions; thus, this work included elements of lesson study (for a brief overview of lesson study see Lewis, Perry, & Murata, 2006).

Data Analysis Plan

Questions were answered with hierarchical linear modeling (*HLM*; Raudenbush, Bryk, Cheong, & Congdon, 2000). We computed a two-level *HLM*, with Level 1 being child and Level 2 being school, to compare pretest-adjusted posttest mathematics achievement between the Building Blocks and control groups.

RESULTS

The results are organized into six categories: (a) compatibility of the treatment groups after random assignment of schools; classroom observations, including (b) observations of the fidelity of implementation of the *Building Blocks* curriculum and (c) observations—using the COEMET—of the quality and quantity of mathematics instruction in all classrooms (this section includes a discussion of the issue of spillover of the intervention into control classrooms); (d) effect of the intervention on the total mathematics achievement score, including a discussion of moderators of that effect; (e) mediational effects of components of the mathematics environment (COEMET) on the total achievement score; and (f) effects of the intervention on individual topics and items.

Compatibility of Groups

The randomized block assignment procedure was reasonably effective in producing equivalent groups (see Table 1). The proportion of SES is close to identical for the two treatment groups. Also important, achievement scores are within about a point of each other on pretest REMA. At the time of the posttest, there was only 5% (n = 70) attrition of children included in the study (18 from control and 52 from Building Blocks). Most of these participants moved out of state. Analyses revealed no significant difference in mean pretest achievement between children who left and children who remained ($F(2) = 2.09$, $p = .148$; $g < .01$). The final sample consisted of 1,305 (927 in Building Blocks, 378 in control) children with complete data on both pretest and posttest.

Classroom Observations: Fidelity

Teachers implemented the *Building Blocks* curriculum with adequate fidelity (see Table 2). That is, on the 5-point Likert scale items, with −2 as strongly disagree and +2 as strongly agree, the mode was 1 in both fall and spring, and the mean was 0.77 in fall and 0.86 in spring. Less than 15% of teachers had an average below 0.50 (about 6% were negative). This result is similar to that observed in previous research with the same instrument and curriculum. For example, the mean in one study was 3.0, equal to Agree in a smaller-scale study using a Likert scale with responses ranging from 1 to 4 (Clements & Sarama, 2008a). (Considering the slight difference between the two results, note that any inclusion of the “neutral” option [0] probably results in means closer to 0.)

Classroom Observations: Quality and Quantity of Mathematics in All Classrooms

The COEMET. Table 2 summarizes data from the COEMET’s three main sections, Classroom Culture, Specific Math Activity(ies) (SMA), and Other Classroom Elements. We used the average of the two time points as our primary scores for analyses because (a) data from the beginning of pre-K year did not constitute a pretreatment measure given that training occurred also in the previous year, so averaging was warranted, and (b) the most complete data were available for these two sets; therefore, we decided that the averaging of these two observations presented the most complete and accurate picture of classroom practice.

ANOVAs revealed Building Blocks classes had higher scores than the control classes on the Classroom Culture subscale ($g = 1.23$, $p < .001$), SMA subscale ($g = 0.78$, $p = .005$), total number of mathematics activities observed in SMAs ($g = 1.02$, $p < .001$), and the number of computers turned on and working for students to use ($g = 0.90$, $p < 001$).

The substantial difference in the number of mathematics activities observed in

SMAs raised the issue of whether this variable was a proxy for total time allocated to mathematics activities. To differentiate and compare these variables, we calculated a new variable that was operationalized as the total number of minutes children experienced mathematics during the visit. The mean time on task was 27.15 (SD = 12.34) minutes for control, and 32.51 (SD = 15.45) minutes for Building Blocks, which was not significantly different (t = 1.70, p = .10).

Spillover into control classrooms. Even with a limited number of observations, data revealed that a small number of control teachers became familiar with specific *Building Blocks* activities before or during data collection. Teachers in the first district who had been part of previous *Building Blocks* curriculum research projects were excited about the curriculum and held summer training sessions for other pre-K teachers who were interested, including control teachers. These control teachers implemented at least some of the *Building Blocks* activities, and used some of the *Building Blocks* materials and activities (e.g., those involving the counting wand were observed).

Similarly, most of the control teachers in the second district had been exposed to the experimental curriculum. After in-service sessions conducted by the developers a few years before this project, this district had embraced the *DLM* curriculum, in that teacher in-service sessions followed these early presentations as teachers shared these activities with each other. As a result, 6 of the 10 control teachers had access to *DLM* activities. One teacher learned about some of the *DLM* activities through her school's mathematics coach. One control teacher had a copy of the *DLM* curriculum open on her desk during a classroom visit by an assessor. Another teacher in the same school mentioned that because her school had been randomly selected as a control school, school administrators had purchased *Building Blocks* materials for her classroom. Two other teachers, in a different school, were using copies of the *DLM* activities, according to their principal. At least three teachers combined *Investigations* activities with *DLM* activities. Such evidence of spillover indicates that our empirical analyses of the effectiveness of the intervention, the achieved relative strength (Hulleman & Cordray, 2009), will be attenuated and thus will constitute an underestimate of the actual intervention effect.

Children's Mathematics Knowledge: Total Score

Based on the significant differences in classroom-level variables, we compared intraclass correlations (*ICC*s) for two- vs. three-level *HLM* models. The *ICC*s for a three-level unconditional model were Level-2 p = .0319 and Level-3 p = .19. To increase the precision of these estimates, the pre-K pretest was included as a covariate at the child, classroom, and school levels (Hedges & Hedberg, 2007). The *ICC*s for this three-level *HLM* with covariates were Level-2 p = .054; Level-3 p = .225. Low *ICC*s for the classroom level suggest relatively small between-classroom variations unique beyond the variance at the school level. Because this study randomly assigned schools rather than teachers, much of the between-class variations is subsumed within the school level. Further, across research groups, 26 schools had only 1 to 2 teacher participants. For these reasons, the impact of treatment group on child achievement was investigated through a two-level *HLM* model in which classroom data were aggregated to the school level. The *ICC*s for the two-level *HLM* were intercepts only: Level-2 $p = .208$, Level-1 $\sigma^2 = 0.439$; with pretest covariates at the child and school level: Level-2 $p = .253$; Level-1 $\sigma^2 = 0.286$.

Analysis of total test. Table 3 presents the means and standard deviations for the REMA's Rasch T-score for all children for whom full data were collected for both pretest and posttest. A two-level *HLM*, with child at level 1, school at level 2, and treatment group entered at Level 2, showed no significant difference in pretest REMA scores between the control and Building Blocks groups ($\beta = -.011$, $SE = .11$, $df = 40$, $p < .29$). A second two-level *HLM* model was constructed including all relevant moderators and pretest aggregates. The full-model equation for the participant level follows.

$$Y_{ij} = \beta_{0ij} + \beta_{1ij}\ (REMAPRE) + \beta_{2ij}\ (Male) + \beta_{3ij}\ (Male * Treatment) + \beta_{4ij}\ (AA) + \beta_{5ij}\ (AA * Treatment) + \beta_{6ij}\ (WHI) + \beta_{7ij}\ (WHI * Treatment) + \beta_{8ij}\ (HIS) + \beta_{9ij}\ (HIS * Treatment) + \beta_{10ij}\ (IEP) + \beta_{11ij}\ (IEP * Treatmen\ + e_{ij},$$

where Y_{ij} is the REMA posttest of child i in school j, β_{0ij} is the initial status of child i in school j; β_{1ij} is the slope of the REMA pretest (*REMAPRE*) for child i in school j; β_{2ij} is the main effect for the dummy code male or not (*male*); β_{3ij} is the interaction of Male and treatment group; β_{4ij} is the main effect of African American or not (*AA*); β_{5ij} is the interaction of African American and treatment group; β_{6ij} is the main effect for the dummy code White or not (*WHI*); β_{7ij} is the interaction of the dummy code for White and treatment group; β_{8ij} is the main effect for the dummy code Hispanic or not (*HIS*); β_{9ij} is the interaction of the dummy code for Hispanic and treatment group (*HIS*Treatment*); β_{10ij} is the main effect for the dummy code has individualized education plan or not (*IEP*); β_{11ij} is the interaction of *IEP* and treatment group; and e_{ij} is the residual (Level-1 random effect). The school-level model (Level 2) follows.

$$\beta_{0j} = \gamma_{0j} + \gamma_{1j}\ (School\ Aggregate\ pretest) + \gamma_{2j}\ (Treatment) + \gamma_{3j}\ (SES) + \gamma_{4j}\ (SES * Treatment) + \gamma_{5j}\ (LEP) + \gamma_{6j}\ (LEP * Treatment) + \mu_j,$$

where β_{0j} is the mean achievement in school j; γ_{0j} is the intercept associated with Level-1 predictors; γ_{1j} is the slope associated with the REMA pretest aggregated to school j; γ_{2j} is the main effect for the dummy coded Treatment group in school j; γ_{3j} is the main effect for *SES* for school j; γ_{4j} is the interaction of SES and Treatment group in school j; γ_{5j} is the main effect for the proportion students with LEP in school j; γ_{6j} is the interaction of LEP and Treatment group (*LEP * Treatment*) in school j; and μ_j is the Level-2 random effect. All variables were centered on the grand mean except for REMA pretest scores, which were centered

Table 3
Means and Standard Deviations for the Early Mathematics Assessment (Rasch and Classical Scores)

	Max		Experimental		Control	
			Pre $n = 927$	Post $n = 927$	Pre $n = 378$	Post $n = 378$
Rasch (total test) score		Mean	18.54	51.36	19.63	44.01
		SD	11.86	17.25	11.52	17.89
Number		Total	9.47	29.71	10.26	25.53
		SD	8.64	12.72	8.27	13.20
Object Counting and Strategies	62	Total	3.61	11.29	3.83	9.57
		SD	3.63	4.91	3.50	5.05
Verbal Counting	8	Total	1.57	4.50	1.85	4.10
		SD	1.88	1.91	1.95	1.87
Comparing and Sequencing	33	Total	2.98	10.22	3.21	8.67
		SD	3.01	4.75	2.85	4.99
Number Recognition and Subitizing	4	Total	1.26	2.60	1.35	2.35
		SD	0.98	1.03	0.99	1.07
Composition of Number	14	Total	0.02	0.35	0.01	0.26
		SD	0.18	0.78	0.14	0.67
Arithmetic	125	Total	0.03	0.75	0.01	0.58
		SD	0.56	2.35	0.10	2.51
Geometry		Total	7.25	14.34	7.65	12.39
		SD	3.87	3.69	3.68	3.71
Shape Identification	22.1	Total	5.01	9.44	5.31	8.08
		SD	2.92	2.00	2.68	2.22
Composing Shape	6	Total	0.06	0.73	0.08	0.44
		SD	0.25	0.76	0.29	0.67
Representing Shape	4	Total	0.09	0.90	0.11	0.67
		SD	0.35	0.80	0.41	0.78
Comparing Shape	7	Total	2.04	2.93	2.11	2.94
		SD	1.20	1.48	1.20	1.46

Table 3 (*continued*)

	Max		Experimental		Control	
			Pre $n = 927$	Post $n = 927$	Pre $n = 378$	Post $n = 378$
Transformations (turns)	3	Total	0.04	0.34	0.04	0.28
		SD	0.20	0.48	0.20	0.45
Patterning	7	Total	0.65	2.63	0.66	2.23
		SD	0.95	1.41	0.94	1.44
Measurement	23	Total	1.18	4.67	1.07	3.85
		SD	0.96	2.48	0.93	2.34

on the group mean for their level. All interactions were computed on mean-centered transformations of the variables involved. Effect sizes were computed for significant main effects by dividing the individual predictor beta coefficient by the pooled posttest standard deviation.

Table 4 presents the results of this analysis. We found a significant main effect for the comparison of the Building Blocks group to the comparison group. The Building Blocks group significantly outperformed the comparison group ($p < .001$), with an effect size (g) of 0.72. Effect sizes for children's posttest scores, controlling for pretest scores, were 1.15 for the control group and 1.92 for the Building Blocks group (using the formula from Lipsey & Wilson, 2001, p. 49).

To examine possible moderators of the treatment effect, Table 4 also presents the interactions between the intervention groups and several Level-2 predictors. School SES (percent of free/reduced lunch) was not a significant predictor of mathematics achievement ($p = .19$), nor was there a significant interaction between the school SES and treatment group ($p = .27$). Similarly, school LEP also was not a significant predictor of mathematics achievement ($p = .08$), nor was there a significant interaction between school LEP and treatment group ($p = .60$).

We also examined possible interactions between treatment group and child-level moderators on child mathematics performance. There were no significant interactions between treatment and the Level-1 variables of gender ($p = .10$) or IEP status (child has an individualized education plan; $p = .37$). Similarly, there was no significant interaction between treatment and child race/ethnicity except one: African American vs. other groups ($p = .015$). In the control group, African American children averaged lower gains on posttest REMA than other children; in Building Blocks classrooms, African American children averaged higher gains than other children.

Table 4 also displays the results of our final model. The final model consists of only those predictors and interactions that were found to be significant in the full model. The final model for the child level follows.

$$Y_{ij} = \beta_{0ij} + \beta_{1ij}\,(REMAPRE) + \beta_{2ij}\,(AA) + \beta_{3ij}\,(AA * Treatment) + e_{ij},$$

Table 4
HLM *Final Fixed Effects Model Outcomes and Variance Components for REMA Mathematics Posttest Scores, With Pretest Scores as Covariates at Both Levels*

	Unconditional model							
	Coeff	*SE*	*df*	*p*				
Intercept (w/cov)	–1.291***	0.051	40	0.000+				
	Random effect	*SE*	*df*	*p*				
Level 1	0.286	0.535						
Level 2	0.097	0.312	40	0.000+				
	Conditional models							
	Full model				Final model			
	Coeff	*SE*	*df*	*p*	Coeff	*SE*	*df*	*p*
Intercept	–1.311***	0.038	35	0.000+	–1.313***	0.032	39	0.000+
Level 1 (Child)								
Pretest								
Covariate	0.462	0.018	1287	0.000+	0.466**	0.018	1299	0.000+
Gender	–0.051	0.030	1287	0.095				
African American	–0.164*	0.069	1287	0.019	–.203***	0.037	1299	0.000+
White	0.055	0.076	1287	0.469				
Hispanic	–0.003	0.073	1287	0.972				
IEP status	–0.047	0.051	1287	0.365				
Level 2 (School)								
Pretest								
Covariate	0.4602***	0.123	35	0.000+	0.564**	0.095	39	0.000+
SES	-0.004	0.002	35	0.190				

Table 4 (*continued*)

LEP	0.004	0.002	35	0.077				
Treatment group	0.4969***	0.068	35	0.000+	0.511***	0.067	39	0.000+
Interactions								
Gender × Treatment group	0.007	0.065	1287	0.912				
African American × Treatment group	0.353*	0.144	1287	0.015	0.231**	0.078	1299	0.004
Hispanic × Treatment group	0.076	0.154	1287	0.621				
White × Treatment group	0.180	0.159	1287	0.260				
IEP status × Treatment group	0.004	0.111	1287	0.974				
SES × Treatment group	–0.005	0.005	35	0.255				
LEP × Treatment group	0.002	0.004	35	0.595				

	Variance components									
	Random effect	*SD*	*df*	*p*	χ2	Random effect	*SD*	*df*	*p*	χ2
Level 1	0.281	0.530				0.282	0.529			
Level 2	0.026	0.170	35	0.00	136.4	0.031	0.175	39	0.00	154.5

where the coefficients are defined similarly to those in the full model by their parenthetical descriptors. The final model for the school level follows.

$$\beta_{0j} = \gamma_{0j} + \gamma_{1j}\,(\textit{School Aggregate pretest}) + \gamma_{2j}\,(\textit{Treatment}) + \mu_j$$

Estimates for the impacts of variables within the final model are displayed in Table 4. Finally, the random effect intercept reveals significant unexplained variance. A following section discusses the mediational hypotheses, which we tested to determine whether this variance can be partially explained as a function of differences in observations of the quality and quantity of mathematics teaching and environment.

Tests of Mediation

In addition to examining the impact of *Building Blocks* on the REMA child posttest outcomes, we investigated the mediational role of instructional processes. Recall that Building Blocks and control teachers differed significantly on five components of mathematics instruction: number of computers, number of SMAs, the classroom culture score, sum of mean SMA scores, and total time on math. The mediational hypothesis is that these components are influenced by the *Building Blocks* curriculum, and they, in turn, cause changes in the outcome variable. A standard regression analysis was conducted utilizing specifications and conventions for subscripts ($_m$) and ($_y$) to reference the mediator and outcome, respectively (Pituch, Stapleton, & Kang, 2006). The equation for this impact is

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

where M_j and X_j represent the school-level mediator and treatment and a represents the impact of the treatment on the mediator. The intercept and residual for the equation are estimated as $\gamma_{00(m)}$ and $u_{0j(m)}$ respectively. To estimate the impact of each mediator on the outcome of child achievement, a two-level hierarchical model was constructed as

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

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

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

where the effect of the mediator on the outcome, b, is estimated controlling for the effect of treatment, and c' is the direct effect of treatment on the outcome controlling for the mediator. The indirect effect is represented by the cross product (ab) of the a and b unstandardized regression coefficients (Preacher & Hayes, 2008). These estimates are displayed in Table 5.

We utilized the empirical M-test to establish 95% confidence intervals for the ab product, submitting the unstandardized regression coefficients a and b and their standard errors to the PRODCLIN program to determine the significance of indirect effects. The empirical M-test corrects for the potentially nonnormal distribution of the *ab* product, determining significance of a confidence interval based on a critical *z* ratio determined across multiple data simulations (MacKinnon, Lockwood, & Williams, 2004). This test has been found to demonstrate low Type 1 error rates, aiding in our effort to identify only relevant mediators (MacKinnon et al., 2004), and it has more power than other approaches to detect smaller effects (Pituch et al., 2006). In addition, we also submitted the unstandardized regression coefficients and standard errors to the interactive Monte Carlo Method for Assessing Mediation (MCMAM) (Selig & Preacher, 2010) for which a normal sampling distribution of the estimates *a* and *b* is simulated multiple times. The 95% confidence intervals within this test of mediation reflect the average 2.5 and 97.5 percentile values across a simulated distribution of 20,000. For both tests, confidence intervals not including zero indicate that the indirect effect of treatment through the intervening variable on child outcome is significant.

Only three indirect effects (from treatment group through a mediator to child outcome) were significant: number of computers on and working for children ($a_1b_1 = 0.11$), number of SMAs ($a_2b_2 = 0.16$), and classroom culture ($a_3b_3 = 0.15$). The sum of mean SMAs ($a_4b_4 = 0.12$) and the total time on math ($a_5b_5 = 0.06$) demonstrated little impact on the relationship between treatment group and child outcome. Confidence intervals around the *ab* product were similar across both the Empirical M-test and the MCMAM.

Children's Mathematics Knowledge: Learning by Topic

Table 3 also presents the means and standard deviations of children's pretest and posttest scores on the REMA using classical scoring. Differences between treatment groups were checked with simple effect sizes (Cohen's *d*) for each subtest, but the nature of these data (gain scores not analyzed with inferential statistics) requires strict caveats regarding their interpretation. They are useful in describing what mathematical topics the intervention was relatively successful in developing and thus suggest future issues for research and development. Results suggest that the learning gains made by the Building Blocks group relative to the control group on number were, in descending order, items involving object counting and counting strategies, verbal counting, comparing number and sequencing, recognition of number and subitizing, composition of number, and arithmetic word problems.

To present a more detailed picture of children's' learning of these mathematical topics, we briefly describe the characteristics of items on which children in the Building Blocks group substantially outperformed the control children and those with smaller relative gains. For object counting and counting strategies, their relative gains were largest on simple object (set) counting and especially production ("give me six . . .") tasks. They also made similar relative gains on counting larger sets (e.g., 15), cardinality tasks ("I covered the [15] pennies you counted. How many are there under the cover?"), and error recognition tasks. Children in the Building Blocks group were more likely to understand a countable unit (e.g., "How

Table 5
Mediation of the Effect of Treatment Group on Child Posttest REMA Outcome Scores, Controlling for Pretest Through COEMET Components

COEMET component	Point estimate	95% confidence intervals PRODCLIN Lower	PRODCLIN Upper	MCMAM Lower	MCMAM Upper
Classroom Culture subscore	0.1537*	0.0140	0.3226	0.0101	0.3269
SMA subscore	0.1211	–0.0172	0.2846	–0.1818	0.2877
Other Classroom Elements					
Total number of SMAs (*M*)	0.1566*	0.0347	0.3133	0.0327	0.3175
Time on task (min/day)	0.0599	–0.0043	0.1609	–.0007	0.1680
Number of computers working for children	0.1145*	0.0195	0.2473	0.0163	0.2503

many whole pencils?" shown a picture of some whole and some broken pencils), but differences were small. The verbal counting items showed large relative gains on the following items: forward counting, counting backward from 10, and counting forward from numbers other than 1.

On comparing number and sequencing, relative gains in favor of the Building Blocks group were largest on comparing numbers presented verbally (e.g., "Which is biggest, 5 or 6 or 4?") and moderate on comparing small sets (less than 5) and on ordering numerals and sets. On both types of problems, the Building Blocks group increased their use of counting and mental strategies (e.g., the mental number line) more than the control group. The items showing the largest relative gains in favor of the Building Blocks group were matching numerals to sets (1–5), administered within this subtest, but actually a combination of reading and mathematics competencies.

The Building Blocks group gained more than the control group on recognition of number and subitizing (Table 3). The greatest gains were on recognizing sets of moderate size (e.g., 4 and 7) with a positive, but smaller, relative gain on sets of 2 and 10.

The greatest relative learning gains made by the Building Blocks group in the remaining subtests were, in descending order, on identifying shapes and their components (these having the most consistent relative gains of any subtests), composition of shape, representing shape, measurement, patterning, and comparing shapes. Regarding specific items on the shape subtest, the greatest relative gains were on shape identification, identification of sides of shapes, and, to a lesser degree, identification of angles. For squares, students in the Building Blocks group were more accurate on all figures, making the strongest relative gains on avoiding the selection of rhombi without 90° angles as well as selecting exemplars and avoiding distractors that shared some characteristics of squares (i.e., regularity and right angles). Results for triangles, rectangles, and rhombi were similar, with the greatest gains on nonprototypical variants (e.g., obtuse triangles with no horizontal side) and on avoidance of distractors with a visual resemblance to prototypical members of each class.

Two other shape items asked about the components of shapes. The Building Blocks group made large relative gains on both side and angle identification.

The Building Blocks group showed large relative gains on the outline puzzle item assessing shape composition. Children in the Building Blocks group increased more on selecting shapes confidently and accurately rotating shapes into correct orientation before placing them on the paper. They also made substantial gains on representing shape, such as accurately constructing shapes with sticks.

Relative gains were higher for the Building Blocks group on all seven items on patterning but were largest on finishing a pattern with an ABB core and next largest on replacing a missing element in a pattern with an AB core. They also gained more in the ability to abstract the core unit from a pattern.

DISCUSSION

We evaluated the *Building Blocks* intervention's impact in multiple schools in two urban school districts, analyzing the specific mathematics concepts and skills that young children learned. We discuss the results in six categories. First we address the preintervention compatibility of the treatment groups. Second, we examine what classroom observations revealed about teachers' fidelity of implementation of the curriculum. Given the large number of schools and teachers, did teachers' enactment achieve the same fidelity as measured in previous, smaller, studies? Third, we examine observations of instruction in both Building Blocks and control classrooms to assess the impact of the intervention on the quality and quantity of mathematics and to address possible spillover of the intervention into control classrooms. Fourth, we discuss the effect of the intervention on total mathematics achievement scores, including moderators. Fifth, we analyze components of the mathematics environment that mediated the intervention's impact on children's learning. Sixth, we explore topics that were differentially affected by the two treatment conditions.

First, analyses revealed that the randomized block assignment was effective in producing equivalent groups. Second, most teachers implemented the curriculum with acceptable fidelity. The modal category for mentor-rated Likert items was agree. That is, most teachers implemented all aspects of the *Building Blocks* curriculum. This result is similar to that observed in previous research with the same instrument and curriculum (Clements & Sarama, 2008a). This provides evidence that interventions in preschool mathematics education, such as this one, can be successfully implemented on a large scale.

Third, the *Building Blocks* intervention enabled teachers to develop richer classroom environments for mathematics than those of the control classrooms as measured by the COEMET. The Building Blocks classes scored significantly higher than the control classes on four components of this measure: the classroom culture subscore, the specific math activities (SMA) subscore, the number of SMAs, and the number of computers on and working for students to use. These findings substantiate the fidelity of the experimental teachers' implementation of the *Building Blocks* intervention and provide evidence that distinguishes these practices from those of the control classrooms. As the name indicates, the classroom culture subscore assesses teachers' general approach to mathematics education indicated by "environment and interaction" variables, such as responsiveness to children and use of "teachable moments" (capitalizing on spontaneous situations that would benefit from mathematization), as well as "personal attributes of the teacher" variables, including appearing knowledgeable and confident about mathematics and showing enjoyment in, curiosity about, and enthusiasm for teaching mathematics. These variables suggest that the *Building Blocks* intervention successfully altered teachers' beliefs and dispositions beyond specific curriculum practices. Such practices were also positively affected, given that teachers used the computer component of the *Building Blocks* curriculum and engaged their children in a greater number of explicit, targeted mathematics activities. The higher scores for the SMA subscale suggest that the specific mathematics activities that teachers in Building Blocks classrooms conducted were higher quality than those in the control classrooms.

Given the significant difference in the number of SMAs and the widely documented importance of time on task (e.g., Bodovski & Farkas, 2007), the question arose whether the number of SMAs was simply a proxy for total time allocated to mathematics activities. To test that hypothesis, we compared the two variables. The total number of minutes during which children were experiencing mathematics was not significantly different in the two treatment groups, nor was this variable significantly related to gain in mathematics knowledge (and it did not mediate the impact of treatment group on children's posttest knowledge of mathematics). Thus, evidence suggests that, at least for children similar to those in this study, the number of distinct mathematics activities in which they engage is more important than total time on task in supporting their learning of mathematics (cf. Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2005, who similarly found that frequency of number activities correlated with achievement, although time on task was not measured in that study). There are several possible explanations for this finding. For example, it may be that developmental constraints, such as limits on attention, result in diminishing returns in longer activities. A more cogent explanation may be that children learn more from a variety of activities emphasizing the same level of thinking; that is, they may learn concepts more readily from generalizing mathematics structures from different problem situations that require the same mathematical concepts and processes (e.g., mental actions-on-objects) for their solution. Further, such multiple situations may create a greater number of cognitive paths for retrieval of these concepts and processes. A caveat is that both the findings and hypothesizing are post hoc and should be tested in future research.

Compared to previous studies (Clements & Sarama, 2007c, 2008a; Klein, Starkey, Clements, Sarama, & Iyer, 2008; Sarama et al., 2008), the counterfactual condition was not the "practice-as-usual" control condition involving no published mathematics curriculum and little district-wide emphasis on mathematics. Both districts had placed new emphasis on pre-K mathematics and adopted new literacy curricula that included specific mathematics components. In addition, the *Building Blocks* intervention itself, especially the 1st year's "gentle introduction," generated considerable spillover of early mathematics pedagogical practices into control classrooms.

Fourth, despite this spillover, children in the Building Blocks group outperformed those in the control group on the total mathematics test score, with an effect size of 0.72. This substantial effect is less than the effect reported for small-scale research (Clements & Sarama, 2007c), and in between those reported in a moderate-size study comparing the Building Blocks group to a comparison group that received a different preschool mathematics curriculum (effect size, 0.47) and to a business-as-usual control group (effect size, 1.07).

Analyses indicated only a single significant moderator of the treatment effect. At the school level, neither socioeconomic status of the school (percent of free/

reduced lunch) nor limited English proficiency predicted children's mathematical achievement. Moreover, neither socioeconomic status of the school nor degree of limited English proficiency significantly interacted with treatment group. Thus, there was no evidence that the *Building Blocks* curriculum was differentially effective in classes serving low- or mixed-income families (with the caveat that the variance of this variable was small) or with schools that have a higher or lower percentage of children with limited English proficiency. At the child level, neither gender nor IEP status predicted mathematics achievement, nor was either a significant moderator of the treatment effect. In sum, there was no evidence that the *Building Blocks* intervention was differentially effective for girls and boys or for children with or without IEPs. There was evidence that the intervention was differentially effective for only a single ethnic/racial comparison: African American children learned less than other children in the same control classrooms and African American children learned more than other children in the same Building Blocks classrooms. It may be that the *Building Blocks* intervention is particularly effective in ameliorating the negative effects of low expectations for African American children's learning of mathematics (see National Mathematics Advisory Panel, 2008).

Fifth, three scores derived from the COEMET instrument, a measure of the quality and quantity of classroom mathematics, mediated the effect of *Building Blocks*. The total number of computers turned on and working for children, the classroom culture component, and the total number of mathematics activities partially, but significantly, mediated the impact of treatment group on outcome. This suggests that these processes support the growth of mathematics learning in children similar to those in this study. The measured mediation was similar to, but less than, the mediational impact found in previous research using the same instrument (Clements & Sarama, 2008a). The finding involving number of computers suggests that increased computer usage, presumably of mathematics software, could lead to improvements in mathematics scores, consistent with research (Clements & Sarama, 2008b). The impact of classroom culture on mathematics scores is consistent with the literature supporting the connection between academic performance and general features of the classroom, including signs of mathematical activity and teachers who are knowledgeable and enthusiastic about mathematics and who interact with and respond to children frequently (Clarke & Clarke, 2004; Clements & Sarama, 2007b). Finally, the mediational impact of the total number of classroom mathematics activities—but not time on task—is an important distinction, as previously discussed. The importance of these factors, however, should continue to be examined across more time points and utilizing more robust mediation methodologies. More specifically, although this analysis investigated the indirect effect of each mediator separately, it is likely that these component processes interact in the creation of an environment more conducive to the learning of mathematics. An area of important future investigation will be to further test the unique and shared contributions of these processes into the kindergarten year.

Sixth, scores of children in the Building Blocks group increased more than those of the control children on all subtests and almost all individual test items (with the caveat that these are only suggestive descriptive statistics of gain scores). However, these relative gains were not similar across mathematical topics. In the domain of number, relative learning gains made by the Building Blocks group were, in decreasing order, on items involving object counting and counting strategies, verbal counting, comparing number and sequencing, recognition of number and subitizing, composition of number, and arithmetic word problems.

In previous research (Clements & Sarama, 2008a), children in Building Blocks classes made consistent relative gains on most verbal counting items (a subtrajectory within the superordinate counting trajectory, with tasks including forward counting, counting backward from 10, and counting forward from numbers other than 1), simple object (set) counting, and production tasks. The children in this study who used Building Blocks showed consistent relative gains in these areas as well, albeit larger for production than for counting small set tasks. Other tasks that showed the same pattern of results involved more sophisticated counting skills, including counting larger sets, cardinality tasks, error recognition tasks, and counting strategy tasks. Although the *Building Blocks* curriculum was the same as that used in previous research, the professional development sessions and tools (e.g., the BBLT application, see Fig. 1) included more examples of such higher level thinking, gleaned from the previous research projects, which may have helped teachers conduct instructional activities developing those skills.

On comparing number and sequencing, relative gains in favor of the Building Blocks group were largest on comparing numbers presented verbally and moderate on comparing small sets (less than five elements) and on ordering numerals and sets. These results are consistent with previous research (Clements & Sarama, 2008a), although the large relative gains on comparing numbers verbally indicate that children in the present study were more successful in developing a "mental number line," another relatively sophisticated strategy emphasized more in the present study's professional development. Similarly, gains on subitizing were more pronounced on larger sets.

Turning to other topics, relative learning gains made by the Building Blocks group were, in decreasing order, items involving identifying shapes and their components, composition of shape, representing shape, geometric measurement, patterning, and comparing shapes. Identifying shapes and their components showed the most substantial and consistent relative gains of any subtest. The largest gains were on items in which children identified or counted the sides and angles of polygons. The Building Blocks group also made substantial gains on representing shape; for example, ensuring that its "rectangle" had all right angles. *Building Blocks* activities in which children discuss, analyze, build, and draw shapes appear to be particularly effective in developing these competencies (especially, perhaps, relative to a counterfactual curriculum with limited time dedicated to geometric and spatial thinking) (Clements & Sarama, 2007c, 2008a).

Analyses of children's specific choices on shape identification tasks similarly indicate that the *Building Blocks* intervention was particularly effective in helping children attend to attributes of shapes, including the number of sides and properties such as equal side length and angle size (e.g., right angles), basing their decisions on these criteria more than on surface-level visual similarity to a prototype. (An exception was a lack of differences between the groups on distractors that were figures that were not closed, indicating a weakness in the enacted curriculum.) Again, *Building Blocks* activities that emphasize children's justifications for their shape naming and constructions (e.g., "Why is this a rectangle?" "How do you know?") appear to be effective in orienting children to begin to think of shapes in terms of their geometric attributes.

In the rectangle identification task, all groups decreased in their selection of squares as examples of rectangles, supporting the notion of a U-shaped developmental pattern in which children's performances initially decrease as they gain (incomplete) knowledge about properties, which only later is completed and consolidated (with visual recognition competencies), enabling error-free performance (Clements, Swaminathan, Hannibal, & Sarama, 1999). Despite attention to this in professional development, the Building Blocks group, compared to the control group, did not select squares more often as members of the class of rectangles or the class of rhombi. Because children in a small number of classes were successful in learning these relationships, we posit that teacher awareness of these relationships was generally inadequate in lieu of specific curriculum tasks.

The Building Blocks group showed large relative gains composing geometric shapes (consistent with Clements & Sarama, 2008a). Children in the Building Blocks group moved away from trial-and-error strategies and used mental imagery and planning as they had in previous studies (e.g., Clements & Sarama, 2007c). This indicates that more children in the Building Blocks group were developing the Shape Composer level of thinking, in which they can build, maintain, and manipulate shapes mentally (Clements et al., 2004).

Shape comparison items showed little gain from any group. Future research should investigate whether (a) such competencies are important at this level and deserve more instructional attention and, if so, (b) what type of instructional task is effective.

The Building Blocks group gained more on geometric measurement items than the control group, but differences were small. Again, if future research determines that more growth is important to young children's learning, more emphasis on measurement and possibly more effective instructional tasks may be warranted.

The Building Blocks group showed large relative gains on patterning. The curriculum emphasizes that the core unit of a pattern and the repetition of this core are the defining attributes of linear sequential patterns. These ideas were effectively appropriated and used by these young children.

In summary, results are generally consistent with previous research, with one additional generalization. The topics on which children in the Building Blocks group made the largest gains relative to the control group, such as shape, shape composition, counting, and comparing number, were those in which research supported the construction of accurate, elaborated learning trajectories at the level of the children's mathematical development (i.e., compared to comparing shape, Clements & Sarama, 2009; Sarama & Clements, 2009). Future research should examine the specific role that the learning trajectories play to ascertain, among the many issues, whether higher quality learning trajectories account for differential gains in learning, or other factors are more important.

IMPLICATIONS

Children, especially those from low-resource communities, need more mathematics education in preschool (Bodovski & Farkas, 2007; Clements & Sarama, 2009; Sarama & Clements, 2009). Evidence from both educational (National Research Council, 2009; Paris, Morrison, & Miller, 2006) and economic (Carneiro & Heckman, 2003) research suggests that early education is the most important period in which to invest resources. The present study provides additional empirical support for the hypothesis that the *Building Blocks* curriculum, as implemented here, helps teachers provide more and better mathematics for their preschoolers.

Further, spillover and compensatory effects, such as those documented in the present study, may reduce measured differences between control and treatment groups (Baker, 2007), indicating that the treatment effects as determined in this study may constitute an underestimate of that benefit, even in large-scale implementations.

A specific policy implication derives from the finding that teacher quality affects children's learning more in low-SES than in high-SES schools, with larger effects on mathematics than reading achievement (Nye, Konstantopoulos, & Hedges, 2004). Thus, curricular-based interventions such the *Building Blocks* mathematics intervention may be especially useful in low-SES schools such as those in this study (cf. Preschool Curriculum Evaluation Research Consortium, 2008). (All schools in this study served a large majority of low-SES children, so the lack of interaction between treatment group and percentage of low-SES children was expected. Results do support the hypothesis that the *Building Blocks* intervention was effective for all groups of children in these schools.) This intervention focuses substantial effort on working with teachers to develop their understanding and effective use of all the tools of the curriculum.

The results provide empirical support for the effectiveness of mathematical learning trajectories as a base for both curriculum and teacher training that engenders shared, systematic practice. It also argues, in contrast to those who champion an individual teacher's idiosyncratic interpretation and implementation of curriculum, that such systematic practice is more effective and amenable to scientifically based improvement than private, idiosyncratic practice (Raudenbush, 2009). This is not to say that teachers could or should implement curricula in routinized ways and certainly not that they should deliver "scripted" curriculum with little or no interpretation. Indeed, such an approach would stand in contraposition to the use of hypothetical learning trajectories in the service of formative assessment. Instead, we propound the following three related points. First, although teachers do interpret curriculum and must be sufficiently knowledgeable and competent to implement it in their classroom context, focusing on the shared scientific base and common goals, such as developmental progressions and instructional tasks in learning trajectories, is a more effective and efficient way to improve education for children as opposed to focusing primarily on teachers' autonomously inventing individual curricula (Raudenbush, 2009). Second, as an additional educational CRT with positive results, this study provides evidence against the precept voiced by some curriculum developers and researchers that experimental studies are inappropriate because different schools and teachers implement curricula in ways so varied that no pattern of results can be found. Third, such scientifically grounded shared practice is, somewhat paradoxically, more likely to generate creative contributions. This is so because they will constitute modifications of effective practice that is already shared, and thus understood, and more easily adopted, and that in turn will be accessible to discussion and further scientific investigation.

In other words, one can agree with William James that "a science only lays down lines within which the rules of the art must fall, laws which the follower of the art must not transgress; but what particular thing he shall positively do within those lines is left exclusively to his own genius . . . many diverse methods of teaching may equally well agree with psychological laws" (James, 1892/1958, p. 24) and yet emphasize that (a) the science of learning and instruction continues to lay down increasingly specific and useful guidelines and (b) teachers and developers who work explicitly within those guidelines and describe the relation between their creative acts and those guidelines will contribute more than those who do not to the present and future practice of education. Idiosyncratic "creativity" that does not build on extant science and learning and instruction is less likely to serve either the profession or the classroom's students.

Addressing similar issues, some researchers have claimed that fidelity to a curriculum contributes to the deprofessionalization of teachers (McClain, Zhao, Visnovska, & Bowen, 2009). Although we agree that a textbook should be a tool, and children's thinking at the core of teachers' practice, our research indicates that such claims are often built upon false dichotomies. Does the published material or the teacher "design the curriculum"? Rather than one or the other, both play critical roles. Most teachers, especially those in early childhood, have limited time and knowledge of mathematics and mathematics education research (Sarama, 2002; Sarama & DiBiase, 2004) required to plan, research, and write truly research-based curricula (as defined in Clements, 2007). However, to enact such a curriculum, they need to understand the mathematics (Ball, Thames, & Phelps, 2008), children's thinking about and learning of that mathematics, and how to design instructional experiences for children at different levels of thinking (Clements & Sarama, 2009; Sarama & Clements, 2009). That is why curricula with a core of scientifically based learning trajectories provide research guidelines within which teachers can be effective professionals. As argued previously, such guidelines are the mark of a profession. Medical doctors are professionals, not "deprofessionalized," when they follow guidelines of scientific knowledge (cf. Raudenbush, 2009). The existence of claims of "deprofessionalization" may stem from several causes. Some institutions do implement fidelity checks that demand mainly compliance to scripts in textbooks, which we believe is a misapplication of the scope and use of scientific knowledge. In other cases, criticism of fidelity may emerge from teachers' impressions that their previous practice is being eliminated and replaced (rather than built upon and improved). Inferring that one's life work is perceived by others to be "faulty" may lead to a dichotomization in which the "fidelity police" are seen as suggesting developmentally inappropriate instruction, so that one can continue comfortably to believe that one has always been engaged in the best professional practice.

Of course, teachers' expertise varies widely, and many have sufficient knowledge of the science and art of teaching to design curriculum. However, in successful, professional settings such as Japan's lesson study, it is a wider community that contributes over a long period of time to modify a single lesson (Lewis et al., 2006; Sowder, 2007). How many teachers can reach such heights, alone, day after day?

In summary, this study adds to a growing number of evaluations suggesting that learning trajectories are useful pedagogical, as well as theoretical, constructs (Baroody, Cibulskis, Lai, & Li, 2004; Clements & Sarama, 2004c, 2008a; Fuson, Carroll, & Drueck, 2000; National Research Council, 2009; Sarama & Clements, 2009; Simon, 1995; Smith, Wiser, Anderson, & Krajcik, 2006). We designed *Building Blocks*' learning trajectories based on analyses of how curriculum materials might serve professional development and professional practice (e.g., Ball & Cohen, 1996; see also the subsequently published Davis & Krajcik, 2005). Both curriculum and practice, that is, implementation, are important, as any curriculum's trajectories are of necessity hypothetical learning trajectories (Simon, 1995), and results of this and previous studies (Clements & Sarama, 2008a) substantiate that teachers' instantiations significantly affect the quality and effectiveness of the intervention. The *Building Blocks* learning trajectories involve a mathematical goal, developmental progressions, and instructional activities (Clements & Sarama, 2007a, 2009; Sarama & Clements, 2009). They are designed to develop teachers' knowledge and pedagogical use of (a) content, by explicating mathematical concepts, principles, and processes required to achieve the mathematical goal; (b) developmental progressions in learning that content, thus encouraging interpretation of children's behaviors and resulting formative assessment (cf. Foorman et al., 2007), as well as a longitudinal view toward the achievement of the goal; and (c) instructional activities, designed to teach children the content and processes (e.g., mental actions-on-objects) necessary to realize each level of thinking in those progressions (this knowledge includes the rationale for the instructional design of each; e.g., why certain length sticks are provided to children with the challenge to build specific shapes). This triad of competencies restrains "lethal mutations" (Brown & Campione, 1996) of instructional activities while encouraging productive adaptations (note that discussing examples of each is one professional development activity).

A caveat is that the roles of each component of the intervention in supporting students' learning, such as the multiple parts of the pre-K curriculum (Clements & Sarama, 2007a) and the professional development sessions, including the Building Blocks Learning Trajectories (BBLT) Internet application as a separate entity, cannot be independently determined. These remain questions for future research. Similarly, we cannot disentangle the contribution of teachers' knowledge and use of the various aspects of learning trajectories. For example, which of the following makes a measurable contribution: greater knowledge of mathematics per se (understanding the goal), knowledge of children's level of thinking, knowledge of instructional tasks, or connections between levels and tasks? Effective formative assessment requires additional knowledge and skill, including knowledge of assessment strategies and how to interpret assessment results and how to use these interpretations to plan instruction for individual or (dynamically constituted) small groups. This study does substantiate the conclusions of previous research that considerable professional development may be necessary to achieve a high-quality implementation of that curriculum. We provided approximately 75 hours of out-of-class teacher training as well as hours of mentoring in the classroom, which is substantially more than offered to most teachers, only 6% of whom participate in mathematics professional development for more than 24 hours over 1 year (Borman et al., 2007). A total of 50 to 70 hours of professional development is consistent with previous research documenting what is necessary to achieve measurable effectiveness (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007).

Educational environments that focus on conceptual understanding and encourage students to develop, discuss, and use strategies for solving challenging problems appear to have similar outcomes. Students learning in those environments do as well on tests of basic number and computational skills and outperform conventional curricula on assessments of thinking, reasoning, problem solving and conceptual understanding (e.g., Boaler & Staples, 2008; Senk & Thompson, 2003; Tarr et al., 2008). Results of this study extend this finding to preschool curricula, given that children in the Building Blocks group appeared to perform as well or better than children in the control group on straightforward items in number and performed substantially better on the most demanding items.

To realize such benefits, it may be necessary to provide support that addresses instruction directly, as did this study's intervention. That is, research suggests that the more proximally linked resources are to classroom instruction, the more likely they are to affect student achievement (National Mathematics Advisory Panel, 2008; Raudenbush, 2009). A large meta-analysis identified domain-specific learning activities as having the strongest impact on cognitive outcomes (Seidel & Shavelson, 2007). The *Building Blocks* intervention addressed specific mathematics and pedagogy for children in the teachers' classrooms. The *Building Blocks* learning trajectories constituted the core of both the specific curriculum the teachers taught and the professional development in which teachers engaged. In this way, all the resources addressed teachers' specific educational problems and solutions. Such targeted professional development positively affects student learning, compared to years of experience or general teacher education, which explained only a small amount of variance in teacher effects (less than 5% and near 0 for early mathematics, Nye et al., 2004).

The results also provide additional support for our Curriculum Research Framework (CRF) (Clements, 2007). Previous studies documented the development of the basic structures and learning trajectories and formative evaluation phases (Clements & Sarama, 2004a; Sarama, 2004). Later studies provided summative evaluations (Clements & Sarama, 2007c, 2008a); however, these studies involved a moderate number of volunteer teachers located in proximity to the researchers. The present study provides a large-scale, multisite summative evaluation with nonvolunteers, completing the tenth and final phase of the CRF and thus providing evidence with direct policy implications. A caveat is that evaluations by developers have been found to have an average effect size of 0.16 greater than those conducted by nondevelopers (Borman, 2007). Borman points out that this may reveal biases, either of the developers or of nondevelopers with a grudge against innovations; however, it is just as likely that the professional development and implementations are superior in the former case. Regardless, our CRF (Clements, 2007) requires that evaluations be confirmed by researchers unrelated to the developers of the curriculum (as noted by Darling-Hammond & Snyder, 1992), with attention given to issues of adoption and diffusion of the curriculum (Fishman, Marx, Blumenfeld, Krajcik, & Soloway, 2004; Rogers, 2003; Zaritsky, Kelly, Flowers, Rogers, & O'Neill, 2003). Such evaluations are needed.

REFERENCES

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

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

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. *Journal of Teacher Education, 59*, 389–407.

Baroody, A. J., Cibulskis, M., Lai, M.-L., & Li, X. (2004). Comments on the use of learning trajectories in curriculum development and research. *Mathematical Thinking and Learning, 6*, 227–260.

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

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *Teachers College Record, 110*, 608–645.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal, 108*, 115–130.

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

Bredekamp, S., Morrow, L. M., & Pikulski, J. J. (2006). *Houghton Mifflin Pre-K: Where bright futures begin!* Boston, MA: Houghton Mifflin.

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

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innova-*

tions in learning: New environments for education (pp. 289–325). Mahwah, NJ: Erlbaum.

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

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

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

Christian, K., Morrison, F. J., Frazier, J. A., & Massetti, G. (2000). Specificity in the nature and timing of cognitive growth in kindergarten and first grade. *Journal of Cognition and Development, 1*, 429–448.

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 of the National Council of Teachers of Mathematics (NCTM) (pp. 67–81). Reston, VA: NCTM.

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

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

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

Clements, D. H., & Sarama, J. (Eds.). (2004c). Hypothetical learning trajectories [Special issue]. *Mathematical Thinking and Learning, 6*(2).

Clements, D. H., & Sarama, J. (2007a). *Building blocks—SRA real math, 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* (Vol. 1, pp. 461–555). Charlotte, NC: 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., & Sarama, J. (2008a). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*, 443–494.

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

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

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

Clements, D. H., Swaminathan, S., Hannibal, M. A. Z., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education, 30*, 192–212.

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.

Cobb, P., McClain, K., Lamberg, T. D., & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher, 32*(6), 13–24.

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

Darling-Hammond, L., & Snyder, J. (1992). Curriculum studies and the traditions of inquiry: The scientific tradition. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 41–78). New York: Macmillan.

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3–14.

Denton, K., & West, J. (2002). Children's reading and mathematics achievement in kindergarten and first grade. Retrieved January 14, 2003, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002125

Economopoulos, K., Murray, M., Eston, R., & Kliman, M. (2008). *Investigations in number, data, and space: Kindergarten* (2nd ed.). Upper Saddle River, NJ: Pearson Scott Foresman.

Fishman, B., Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *The Journal of the Learning Sciences, 13,* 43–76.

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

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

Fuson, K. C., Carroll, W. M., & Drueck, J. V. (2000). Achievement results for second and third graders using the *Standards*-based curriculum *Everyday mathematics. Journal for Research in Mathematics Education, 31*, 277–295.

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.

Heckman, J. J. (2003). Policies to foster human capital. *Research in Economics, 54*, 3–56.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*, 60–87.

Horizon Research, Inc. (2001). *Classroom observation protocol*. Chapel Hill, NC: Author. Retrieved October 2, 2002, from http://www.horizon-research.com/about/facilities.php

Houghton Mifflin Harcourt. (n.d.). Houghton Mifflin Harcourt School Publishers. Retrieved December 15, 2009, from http://www.hmhschool.com/School/index.html

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2,* 88–110.

James, W. (1958). *Talks to teachers on psychology: And to students on some of life's ideas.* New York: Norton. (Original work published 1892)

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness, 1*, 155–178.

Krueger, A. B. (2003). Inequality: Too much of a good thing. In J. J. Heckman & A. B. Krueger (Eds.), *Inequality in America: What role for human capital policies?* (pp. 1–75). Cambridge, MA: MIT Press.

Lewis, C. C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher, 35*(3), 3–14.

Lipsey, M. W., & Wilson, D. B. (2001). *Applied social research methods series: Vol. 49. Practical meta-analysis* (L. Bickman & D. J. Rog, Series Eds.). Thousand Oaks, CA: Sage.

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

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal, 41*, 115–157.

McClain, K., Zhao, Q., Visnovska, J., & Bowen, E. (2009). Understanding the role of the institutional context in the relationship between teachers and text. In J. T. Remillard, B. A. Herbel-Eisenmann, & G. M. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 56–69). New York: Routledge.

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: Findings from IEA's repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Boston: The International Study

Center, Boston College, Lynch School of Education.

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

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

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

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

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity* (C. T. Cross, T. A. Woods, & H. Schweingruber, Eds.). Washington, DC: The National Academies Press.

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

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

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

Paris, S. G., Morrison, F. J., & Miller, K. F. (2006). Academic pathways from preschool through elementary school. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of research in educational psychology* (2nd ed., pp. 61–85). Mahwah, NJ: Erlbaum.

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

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–889.

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

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

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

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. J, Jr. (2000). *HLM5: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: The Free Press.

Sarama, J. (2002). Listening to teachers: Planning for professional development. *Teaching Children Mathematics, 9*, 36–39.

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: Erlbaum.

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

Sarama, J., & Clements, D. H. (2011). *TEAM—Tools for early assessment in mathematics*. Columbus, OH: McGraw-Hill Education.

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

Sarama, J., Clements, D. H., & 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 eighteenth annual meeting of the North American Chapter of the Inter-*

national Group for the Psychology of Mathematics Education (Vol. 2, pp. 567–572). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

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

Schickedanz, J., Dickinson, D. K., & Charlotte-Mecklenberg Schools. (2005). *Opening the world of learning: A comprehensive literacy program*. Parsippany, NJ: Pearson Early Learning.

Schiller, P., Clements, D. H., Sarama, J., & Lara-Alecio, R. (2003). *DLM early childhood express*. Columbus, OH: SRA/McGraw-Hill.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*, 454–499.

Selig, J. P., & Preacher, K. J. (2010). Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [Online applet]. Retrieved August 2, 2010, from http://www.people.ku.edu/~preacher/medmc/medmc.htm

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

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

Smith, C. L., Wiser, M., Anderson, C. W. A., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic-molecular theory. *Measurement: Interdisciplinary Research & Perspective, 14*(1 & 2), 1–98.

Sowder, J. T. (2007). The mathematical education and development of teachers. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 157–223). Charlote, NC: Information Age.

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

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2005). *The Effective Provision of Pre-School Education [EPPE] project: A longitudinal study funded by the DfEE (1997–2003) Technical Paper #6*. London: EPPE Project, Institute of Education, University of London.

Tarr, J. E., Reys, R. E., Reys, B. J., Chávez, Ó., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*, 247–280.

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

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. (2007). *State and local implementation of the No Child Left Behind Act, Vol. II—Teacher quality under NCLB: Interim report*. Washington, DC: U.S. Department of Education.

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

Weiss, I. R. (2002, April). *Lessons learned about systemic reform of mathematics education*. Presentation at the Research Presession of the eightieth annual meeting of the National Council of Teachers of Mathematics, Las Vegas, NV.

Whitehurst, G. (2009). *Don't forget curriculum*. Washington, DC: Brown Center on Education Policy, The Brookings Institution.

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

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 & 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.

Zaritsky, R., Kelly, A. E., Flowers, W., Rogers, E., & O'Neill, P. (2003). Clinical design sciences: A view from sister design efforts. *Educational Researcher, 32*(1), 32–34.

Authors

Douglas H. Clements, University at Buffalo, State University of New York, Learning and Instruction, 212 Baldy Hall (North Campus), Buffalo, NY 14260; clements@buffalo.edu

Julie Sarama, University at Buffalo, State University of New York, Learning and Instruction, 212 Baldy Hall (North Campus), Buffalo, NY 14260; jsarama@buffalo.edu

Mary Elaine Spitler, University at Buffalo, State University of New York, Learning and Instruction, 212 Baldy Hall (North Campus), Buffalo, NY 14260; mspitler@buffalo.edu

Alissa A. Lange, University at Buffalo, State University of New York, Learning and Instruction, 212 Baldy Hall (North Campus), Buffalo, NY 14260; alange@buffalo.edu

Christopher B. Wolfe, University at Buffalo, State University of New York, Learning and Instruction, 212 Baldy Hall (North Campus), Buffalo, NY 14260; cwolfe1ster@gmail.com

Accepted: August 8, 2010