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

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

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

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

Background and Theoretical Framework

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

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

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

In the second category, Learning Model, developers structure activities in accordance with empirically based models of children's thinking in the targeted subject-matter domain. This phase, (4) Structure According to a Specific Learning Model, is critical to this study; therefore, we will elaborate. We created research-based *learning trajectories* (Simon, 1995), which we define as "descriptions of children's thinking and learning in a specific mathematical domain, and a related, conjectured route through a set of instructional tasks designed to engender those mental processes or actions hypothesized to move children through a developmental progression" (Clements & Sarama, 2004c, p. 83). For example, children's developmental progression for shape composition (Clements, Wilson, & Sarama, 2004; Sarama, Clements, & Vukelic, 1996) advances through levels of trial and error, partial use of geometric attributes, and mental strategies to synthesize shapes into composite shapes. The sequence of instructional tasks requires yields better recall and retention (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Rohrer & Taylor, 2006). Fifth, interweaving may facilitate mutual reinforcement between learning trajectories (Clements & Sarama, 2007b). As a simple example, early learning of subitizing (rapid recognition of the numerosity of small sets) supports the development of a critical (and oft-neglected) level of thinking in the counting trajectory in which children recognize the last counting word indicates how many in the counted set. That is, if children count a group of objects, "One, two, three," and immediately recognize a group as containing three objects via subitizing, their understanding of the cardinality of the last counting word is facilitated. Conversely, the establishment of that level of thinking in counting supports the development of higher levels of subitizing.

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

The present study is the first of several Phase 10 evaluations (Clements, 2007) evaluating the effects of a complete preschool mathematics curriculum on the mathematical knowledge of 4-year-old children, including those attending schools that serve children from low-SES families. Research questions included the following: Can *Building Blocks* be implemented with high fidelity, and does the measure of fidelity predict achievement gains? Does *Building Blocks* have substantial positive effects on the quality of the mathematics environment and teaching? What are the effects of the *Building Blocks* curriculum, as implemented under diverse conditions, on the mathematics achievement of preschoolers? A final, secondary, question was, If these effects are significant, does the increase in the quality of the mathematics environment and teaching mediate the effects on mathematics achievement? The complexity of numerous contexts, compared to the

Table 1

Demographics of Participating Schools

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

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

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

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

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

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

Table 2
Comparison of Curriculam

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

(continued)

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

Measures

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

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

(text continues on p. 461)

Table 3
(continued)

	Comparison				*Building Blocks*			
	Observation			Mean	Observation			Mean
	1	2	3	3 Obs.	1	2	3	3 Obs.
Organization, teaching, approaches, interactions	3.2	3.0	3.2	3.1	3.1	3.2	3.3	3.2
Materials were set up correctly[b]	4.0	4.0	4.0	4.0	3.8	4.0	3.8	3.9
Teacher conducted activity as written	3.7	4.0	3.3	3.7	2.9	3.1	3.5	3.2
Pace was appropriate	3.6	3.4	3.7	3.6	3.6	3.3	3.4	3.4

Table 3

Table 4

Table 4
(continued)

		Control				Comparison				*Building Blocks*			
		Observation			*M*	Observation			*M*	Observation			*M*
Scored Items	Diff.	1	2	3	3 Obs.	1	2	3	3 Obs.	1	2	3	3 Obs.
Enthusiasm for math ideas	45.8	3.1	2.7	2.9	2.9	3.7	3.6	3.6	3.6	3.1	3.4	3.1	3.2
Specific math activities													

Table 4
(continued)

	Control				Comparison				*Building Blocks*			
	Observation			*M*	Observation			*M*	Observation			*M*
Score (Building BloSupportj/Fchs) Diff.’ f -32.2 1837 -2.9 c5966 M OTwt items Trem. 1 2 *M*				3 Obs.	1	2	3	3 Obs.	1	2	3	3 Obs.

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

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

Children's mathematical knowledge. The third instrument measured children's mathematical knowledge and skills. Other instruments were deemed too limited in coverage (e.g., restricted topics, usually only number, and restricted range, such as the Woodcock-Johnson's multiple tasks on numbers 1 to 4). No available instruments avoided these limitations, according to two national panels on preschool assessment (NICHD Forum, Washington, DC, June 2002; CIRCL Forum, Temple University, January 30–31, 2003). Thus, we used the Early Mathematics Assessment (EMA), a measure of preschool children's mathematical knowledge and skills that features two individual interviews of each child of about 10 to 20 minutes, with explicit protocol, coding, and scoring procedures. All sessions are videotaped, and each item is coded for accuracy and, when relevant, for solution strategy used by two trained coders. Any discrepancies are resolved via consultation with the senior researchers. The EMA assesses children's development in a comprehensive set of mathematical topics (see Table 5), rather

***BBLT* provides scalable access to the learning trajectories via descriptions, videos, and commentaries. Each aspect of the learning trajectories—*developmental progressions* of children's thinking and connected *instruction*—are linked to the other. For example, teachers might choose the 

(curriculum) view and see the screen on the left, below. Clicking on a specific activity provides a description. Clicking on**

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

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

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

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

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

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

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

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

Analyses. Factorial repeated measures analyses were conducted on the

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

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

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

Results

Fidelity

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

The repeated-measures ANOVA computed on the Fidelity *T* score was not significant for time (fall, winter, spring), $F(2, 38) = .33$, $p = .73$, $MSE = 47.26$); treatment $F(1, 19) = .07$, $p = .80$, $MSE = 1$ Tting activities with all children, asl relatively more difficult to achieve for this population). The *Building Blocks* group employed a greater number of different activities than the other two groups. The number of computers running mathematics activities was highest in the *Building Blocks* actively engaged in activities was of moderate difficulty. The degree the

Table 7

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

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

Table 7
(continued)

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
No response	2.0	0.0	2.1	0.0	1.0	1.0
Item not administered	0.0	0.0	2.1	0.0	0.0	1.0
3. Identify another's mistake in object counting						

Table 7
(continued)

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
5. Count scrambled arrangement of 15 objects						
Reading order	6.0	2.0	0.0	4.3	3.0	1.0

Table 7
(continued)

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

(continued)

Table 7
(continued)

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

(continued)

Table 7
(continued)

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

(continued)

Table 7
(continued)

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

(continued)

Table 7
(continued)

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
No response	8.91	2.97	9.80	1.96	12.87	6.93
Item not administered	0.99	0.00	0.00	0.00	0.99	0.00

Table 7
(continued)

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
20. Construct rectangle with straws (angle size)						
Incorrect on all	4.95	7.92	5.88	0.00	2.97	0.99

Table 7
(continued)

	Control		Comparison		*Building Blocks*	
Item/Response	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
22. Composing shape (accuracy)						
Turned shapes after placing on puzzle in an attempt to fit	11.88	3.96	1.96	7.84	6.93	1.98
Turned shapes into correct	31.68	48.51	5.88	60.78	12.87	82.18

Table 7
(continued)

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

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

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

Both intervention groups scored higher than the control group, with little difference between them, on object counting and verbal counting strategies, comparing number, and sequencing (Table 5). On object counting, the most consistent relative gains were on simple object counting and production ("Give me six . . .") tasks as well as some sophisticated counting strategy tasks (e.g., "Here are six pennies. There are three more under this cloth. How many are there in all?"). By the posttest, 95% of all children used effective counting strategies for counting five objects in a line. The comparison group used such strategies less at pretest than the other two groups and thus showed more increase in the behavior of correctly touching or pointing to each object. The *Building Blocks* BuildingBlocksl

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

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

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

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

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

Consistent with the *T* score results, the *Building Blocks* group scored higher than the comparison group, which scored higher than the control group, on shape composition (Table 5). Again, the comparison group’s greater gains than the control group resulted from an increase in partially correct solutions. The *Building Blocks* group increased more than the other two groups in completely correct solutions (Table 7, Item 21). Similarly, the *Building Blocks*

observable mathematics in the environment and the quality of teachers' pedagogical strategies, and results suggest that this increase partially accounted

particularizes theories of curriculum research (Clements, 2002, 2007). An implication is that such synthesis of curriculum development as a scientific enterprise and mathematics education research may help reduce the separation of research and practice and produce results that are applicable not only to researchers but to practitioners (parents, teachers, and teacher educators), administrators, policy makers, and curriculum and software developers. A caveat is that this study involved a moderate number of mostly volunteer teachers located in proximity to the researchers.

Efficacy of the Learning Trajectories Construct

The larger effects of the *Building Blocks* curriculum than the comparison curriculum support the distinct contribution of research-based learning trajectories. Given that the content coverage of the two were closely matched, significant differences favoring the *Building Blocks* curriculum may be the result of the instructional strategies and the pedagogical content knowledge embedded in its learning trajectories. These results intimate that others will find learning trajectories a useful construct in future research, curriculum development, and professional development efforts. Learning trajectories have several advantages, including that (a) the developmental progressions provide benchmarks for formative assessments, especially useful for children who are low performlearning trajectories, including comparing it to alternatives, to evaluate the effectiveness of the model per se.

Note

This paper was based upon work supported in part by the National Science Foundation

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

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

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

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

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

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.
Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor,

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