Journal for Research in Mathematics Education
2016, Vol. 47, No. 3, 206–232

Improving Mathematics Learning of Kindergarten Students Through Computer-Assisted Instruction

Matthew E. Foster and Jason L. Anthony
University of Texas Health Science Center at Houston

Doug H. Clements and Julie Sarama
University of Denver

Jeffrey M. Williams
University of Texas Health Science Center at Houston

This study evaluated the effects of a mathematics software program, the Building Blocks software suite, on young children's mathematics performance. Participants included 247 Kindergartners from 37 classrooms in 9 schools located in low-income communities. Children within classrooms were randomly assigned to receive 21 weeks of computer-assisted instruction (CAI) in mathematics with Building Blocks or in literacy with Earobics Step 1. Children in the Building Blocks condition evidenced higher posttest scores on tests of numeracy and Applied Problems after controlling for beginning-of-year numeracy scores and classroom nesting. These findings, together with a review of earlier CAI, provide guidance for future work on CAI aiming to improve mathematics performance of children from low-income backgrounds.

Key words: At-risk students; Computer-assisted instruction; Kindergarten; Mathematics; Numeracy

All citizens need a broad understanding of mathematics to function in today's society, but mathematics proficiency rates in the United States are low (Ginsburg, Leinwand, Anstrom, & Pollock, 2005; Kilpatrick, Swafford, & Findell, 2001). International comparisons indicate that children in the United States perform

---

This material is based on research that was supported by the Institute of Education Sciences (IES) under Grant No. R305A080196, and this manuscript was written with the support of IES Grant No. R324B110007, awarded to the University of Texas Health Science Center in Houston. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES. We would like to thank all participants and their families and schools for their cooperation and support. We also greatly appreciate the work of all contributing members of the Children's Learning Institute.

The present study was designed, implemented, and analyzed independent of the authors of the Building Blocks program (Dr. Douglas H. Clements and Dr. Julie Sarama), who contributed substantially to the intellectual property of this manuscript, and the publisher of the Building Blocks program (McGraw-Hill Education), which provided the software suite and support services at greatly reduced costs.

worse in mathematics, and their lagging mathematics development is evident as early as preschool (Cross, Woods, & Schweingruber, 2009; Sarama & Clements, 2009; Starkey, Klein, & Wakeley, 2004). Importantly, children who live in poverty and who are members of linguistic and ethnic minority groups demonstrate significantly lower levels of mathematics achievement than their majority, middle-class peers (Clements & Sarama, 2011; Denton & West, 2002; National Assessment of Educational Progress [NAEP], 2013). Moreover, the achievement gap is wider in the United States than in any other country in the world (Akiba, LeTendre, & Scribner, 2007). Given that early mathematics knowledge is a stronger predictor of later mathematics achievement than even intelligence or memory abilities (Krajewski, 2005) and that children who begin with the lowest achievement levels show the lowest mathematics growth from Kindergarten to the third grade (Bodovski & Farkas, 2007), such achievement gaps in mathematics are pernicious (Claessens, Duncan, & Engel, 2009; Clements & Sarama, 2011; Horn, 2005; National Mathematics Advisory Panel [NMAP], 2008).

In the United States, low socioeconomic status and status in some minority groups are risk factors for low mathematics achievement, which has been attributed to lack of opportunities to learn mathematics (Clements & Sarama, 2009). Moreover, high-quality instructional experiences early in the lives of children can help to improve mathematics achievement and help to prevent or mitigate the development of mathematics learning difficulties (Clements & Sarama, 2011; Cross et al., 2009; Magnuson, Meyers, Ruhm, & Waldfogel, 2004). Benefits derived from effective mathematics instruction provided during the preschool to early elementary school period appear greatest for children from low-income families and whose parents have little education (Case, Griffin, & Kelly, 1999; Clements, Sarama, Spitler, Lange, & Wolfe, 2011; Griffin & Case, 1997). Instructional effects have been shown to continue into the late elementary school and high school years (Brooks-Gunn, 2003; Friedman-Krauss & Barnett, 2013). However, few rigorous evaluations of Kindergarten mathematics programs vetted by the What Works Clearinghouse (WWC) exist (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, n.d.). Of the 40 mathematics programs listed in the WWC, only one pertained to Kindergartners, the DreamBox Learning program. Of the 11 evaluation studies that focused on this computer program, only one met the WWC highest category of evidence, "evidence standards without reservations"; none of the other 10 evaluations met even low evidence standards (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2013). Thus, there is a pressing need for rigorous research on Kindergarten mathematics interventions in the United States, especially those aimed at improving mathematics outcomes for children from low-income and minority groups, which is the aim of the present study.

The Need for Research on Computer-Assisted Instruction

Computer-assisted instruction (CAI) refers to computer software programs that help students learn mathematics or provide an opportunity for students to learn and apply mathematics concepts and skills (Harskamp, 2015); it is most often used as a supplement to children's classroom instruction (Slavin & Lake, 2008). CAI is a potentially attractive means for supporting mathematics learning due to its ease of implementation, standardized scope and sequence of curriculum, and suitability for individualized instruction through regular monitoring of children's progress coupled with adaptive instruction (Anthony, 2015; Clements & Sarama, in press). However, educators' questions and concerns about developmental appropriateness, logistics of implementation, and compatibility with core curricula as well as concerns about effectiveness hamper widespread use of CAI during the early school years. Reviews of the scientific literature have generally concluded that when used appropriately, CAI can provide substantial benefits for children's mathematics learning (Baroody, Eiland, Pupura, & Reid, 2013; Clements & Sarama, 2010, in press; Cross et al., 2009; Lentz, Seo, & Gruner, 2014; Li & Ma, 2010; Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009; Slavin & Lake, 2008). Moreover, research findings have suggested that children in preschool (Fletcher-Flinn & Gravatt, 1995) and primary grades, especially those in compensatory educational settings (Lavin & Sanders, 1983; Niemiec & Walberg, 1985; Ragosta, Holland, & Jamison, 1982), have made the largest gains in mathematics from CAI.

One review (NMAP, 2008) indicated that CAI applications that are well designed and implemented can have a positive impact on mathematics performance, and recent studies support this conclusion (Harskamp, 2015; Moradmand, Datta, & Oakley, 2013; Nusir, Alsmadi, Al-Kabi, & Sharadgah, 2013). A more recent meta-analysis of rigorous studies similarly concluded that there are positive effects of CAI in mathematics when used as a supplement to children's daily classroom instruction (Cheung & Slavin, 2013). Another meta-analysis of studies examining the use of CAI for early mathematics learning found a moderate effect size (Harskamp, 2015), whereas still another meta-analysis found positive effects for the use of technological manipulatives (Moyer-Packenham & Westenskow, 2013). Therefore, CAI represents a viable medium to deliver supplemental mathematics instruction.

Although results from meta-analyses suggest that supplemental use of educational software has an overall positive effect, one large and noteworthy randomized control trial indicated virtually no effect for computer use with older students and several varieties of software (Dynarski et al., 2007). Further, the work of Cuban (2001) indicated that computers can often be "oversold and underused" in early childhood education. Cuban (1993) argued that the efficacy of computer use in the classroom is too frequently doomed to failure because teachers are reluctant to use the technology, particularly because technology is thought to take time away from teacher–student interactions. Others have argued that CAI may focus only on basic skills for students from low-income and minority backgrounds, further limiting their opportunities to learn mathematical reasoning (Kitchen & Berk, 2016). Given these concerns and conflicting empirical results, it is important to examine the effectiveness of educational software programs aimed at improving children's mathematical competencies.

The Building Blocks Mathematics Program

The Building Blocks mathematics program is comprehensive in that it includes a teacher's edition, assessment and resource guides, manipulatives, "big books," and a software suite (Clements & Sarama, 2013). The program, including the Building Blocks software suite, aims to develop understanding and skill fluency in the domains of numeracy and geometry. Three mathematical subthemes are woven throughout both domains: patterns, data, and sorting and sequencing. Although initially developed before the Common Core State Standards for Mathematics (CCSSM; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), Building Blocks content coverage is aligned closely with the CCSSM, which specifies grade-level standards for the areas of counting, arithmetic operations, place value, measurement, and geometry. The Building Blocks software suite targets developmental learning trajectories that span from preschool to third grade. Its more than 200 activities are organized into topical learning trajectories that were designed based on a comprehensive Curriculum Research Framework (Clements & Sarama, 2007; Clements, 2007) and a specific model, consistent within that framework, that details the development of scientifically based software (Clements & Battista, 2000). Thus, the software suite is research based in several fundamental ways. Research-based computer tools stand at the base, providing computer analogs to critical mathematical ideas and processes. These are implemented with activities and a management system that guides children through fine-grained, research-based learning trajectories. These activities are designed to connect children's informal knowledge to more formal school mathematics. The result is a software package that is motivating for children but is also comprehensive in that it includes both exploratory environments that include specific tasks and guidance, building concepts and well-managed practice in building skills, and a full range of mathematical activities.

The design process for the curriculum and the software was based on the assumption that both can and should have an explicit theoretical and empirical foundation. It also should interact with the ongoing development of theory and research—reaching toward the ideal of testing a theory by testing the software or the curriculum in which it is embedded. The model includes specification of mathematical ideas (computer objects or manipulatives) and processes or skills (software tools or actions) and extensive field testing from inception through large summative evaluation studies (Clements & Battista, 2000; Clements & Sarama, 2007, 2008; Sarama & Clements, 2009). Thus, this study represents not just an evaluation of one particular suite of technology-based activities but a rigorous test of the efficacy of software designed on scientific principles.

A series of empirical studies have supported the effectiveness of the full Building Blocks curriculum, which includes the classroom components along with the integrated software suite (e.g., Clements & Sarama, 2007, 2008; Sarama & Clements, 2009). One evaluation of the Building Blocks program demonstrated the curriculum's effectiveness with preschool children from low-income backgrounds (Clements & Sarama, 2007). Using a proximal measure closely aligned with the curriculum's hypothesized learning trajectories, Clements and Sarama (2007) demonstrated that the mathematics achievement of children in two experimental classrooms who used the Building Blocks program increased significantly more than that of children in the two comparison classrooms, who used the school's existing curriculum. Thus, achievement gains of the experimental group indicated that the Building Blocks program was effective under controlled conditions.

Two additional studies examined Building Blocks on a larger scale. The first study (Clements & Sarama, 2008) involved 36 teachers randomly assigned to one of three conditions: (a) Building Blocks, (b) another pre-Kindergarten mathematics curriculum (with equal professional development and resources), and (c) control (or business as usual). Estimates of effect size (Cohen's d) indicated that the mean gain for early numeracy and geometry skills for children in the Building Blocks group was significantly greater than that of children in the comparison group ($d = 0.47$) and control group ($d = 1.07$). The study therefore demonstrated that the comprehensive Building Blocks curriculum was effective—that is, it contributed to children's mathematics learning when implemented outside of controlled conditions by teachers of state-funded pre-K and Head Start classrooms. A second randomized control trial (Clements et al., 2011) involved 106 teachers and 1,305 children attending 42 schools serving children from low-income backgrounds in two states. The results corroborated those of the first study and demonstrated that teachers effectively implemented the Building Blocks curriculum and that students in classrooms where Building Blocks was implemented learned more mathematics than students in the comparison classrooms (Hedge's $g = 0.72$).

Classroom observational data gathered as part of the three studies described above indicated that use of the computer software suite was one of the factors that was most highly correlated with children's gains in mathematics achievement. For example, in the study of 36 classrooms (Clements & Sarama, 2008), the number of computers running the Building Blocks software was one of the variables that most differentiated the treatment groups andcorrelated with children's gains in mathematics. In the study of 106 teachers (Clements et al., 2011), the classroom observation data indicated that the number of computers running the Building Blocks Software had one of the three highest relationships with child outcomes, mediating the effect of treatment on student gains in mathematics while accounting for the effects of two other mediators: structured mathematics activities and classroom culture. In summary, a number of studies have demonstrated that the Building Blocks curriculum is efficacious and effective. There is also correlational evidence suggesting that experimental classrooms in which more computers were running the CAI component of the Building Blocks program tended to have students who learned more mathematics than did students in experimental classrooms in which fewer computers were running the CAI component. However, random assignment to experimental conditions, with one condition that included use of the software and another condition that did not, was not employed in the studies above. The aforementioned studies, therefore, do not allow for causal inferences concerning the impact of CAI via the Building Blocks software, either within or outside of the context of the larger Building Blocks curriculum.

Purpose of the Current Study

Poor mathematics achievement of students from low-income and ethnic minority backgrounds is acknowledged as a problem of national significance (Jordan, Kaplan, Ramineni, & Locuniak, 2009). The extant literature, however, includes few rigorous evaluations of mathematics curricula (Rutherford et al., 2014), especially of CAI programs for Kindergarteners. Consequently, little is known with regard to the "ingredients" needed for specific instructional methods or programs to be most effective at improving the mathematical competencies of young children, especially those at risk for school failure (Jordan et al., 2009). Therefore, there is a pressing need to identify effective mathematics programs for Kindergartners from low-income and minority backgrounds. Our interest in evaluating the Building Blocks software suite in these populations is motivated by the fact that for decades, children from low-income and some minority backgrounds have demonstrated substantially lower levels of mathematics achievement than their majority, middle-class peers (NAEP, 2013). It is also noteworthy that the Building Blocks software suite was developed for children between the ages of 4 to 9 years, when prevention efforts are likely to prove most beneficial (Clements, Baroody, & Sarama, 2013; Clements et al., 2011). To this end, the present randomized control study addressed the following research question: When used as a supplement to typical classroom instruction, does use of the Building Blocks software lead to improvements in mathematics achievement as measured by proximal and distal measures of children's mathematics performance? It was hypothesized that low-income minority children who received supplemental CAI via the Building Blocks software would demonstrate higher levels of mathematics achievement than children who did not receive CAI via the Building Blocks software.

Method

This randomized control study included three annual cohorts of students. All students received some form of CAI throughout most of their Kindergarten school year. To determine the effectiveness of the Building Blocks software program, we used multilevel modeling to examine the effect of the experimental condition on Kindergarten mathematics outcomes while controlling for beginning-of-the-year mathematics performance and classroom-level differences. To determine treatment specificity, we examined the effect of the Building Blocks software on children's vocabulary scores.

Participants and Contexts

Schools. Nine schools participated over the course of the 3-year project. These schools were located in a large urban school district in Texas and were chosen because they served a large population of low-income children. Indeed, the percentage of students eligible for free or reduced lunch programs at each school ranged from 70% to 98% (*M* = 92.2, *SD* = 8.6). Participating schools primarily served African American and Hispanic children.

Teachers. Across three annual cohorts, 27 Kindergarten teachers participated. All teachers were certified by the state to teach in the public school system. All teachers were female, college-educated, and native or fluent speakers of English; two were also fluent in Spanish. Data about teachers' ethnicity were reported for 22 of the 27 teachers: 15 were African American, five were Caucasian, and two were Hispanic. Because three annual cohorts of Kindergarten children participated and because some teachers were assigned to the same grade in subsequent years, a few teachers participated in more than 1 year. Specifically, two teachers participated all 3 years, four teachers participated in 2 years, and 23 teachers participated in 1 year, resulting in a total sample size of 37 Kindergarten classrooms.

Classrooms. All of the children in the 37 classrooms participating in the study attended full-day Kindergarten. The average class size was 21 students. On average, 95% of the students were native speakers of English, and 5% were native speakers of Spanish who were learning English as a second language. Teachers in the majority of classrooms (60%) reported that 100% of their mathematics instruction was provided in English, whereas teachers in the remaining classrooms (40%) reported providing mathematics instruction in Spanish to varying degrees. More specifically, teachers in 16% of classrooms reported that 50% to 85% of their mathematics instruction was provided in Spanish. The remaining 24% reported that 90% to 100% of their mathematics instruction was provided in Spanish. Teachers reported that daily mathematics instruction ranged between 30 and 125 minutes (*M* = 78.55 minutes, *SD* = 20.17). When asked to report their goals for classroom instruction and individual teaching opportunities on a 5-point Likert scale that reflected relative proportions of emphasis on understanding mathematical concepts versus practice (without an emphasis on understanding mathematical concepts), teachers in 46% of the classrooms reported equally emphasizing conceptual understanding and practice. Teachers in the remaining classrooms reported emphasizing concept building with some practice (39%), emphasizing conceptual understanding (11%), and emphasizing practice (4%). No teachers reported emphasizing practice with some emphasis on conceptual understanding.

Participating classrooms reflected commonplace instructional practices in Kindergarten that balanced discovery learning with explicit instruction. In almost all classrooms (98%), instruction followed a version of the enVisionMath mathematics program for Kindergarteners. That program included daily lesson plans organized into focused topics aimed at developing students' conceptual understanding through practice and problem-solving activities. All teachers were allowed to alter their pace through the program and to supplement their mathematics instruction with materials and activities that were not part of the enVisionMath program. In fact, of higher priority than following the program was adherence to the state curriculum standards, Texas Essential Knowledge and Skills (TEKS),[1] and adherence to the school district's written expectations for student learning, which were very closely aligned with TEKS. In doing so, teachers may have used materials in addition to their classroom mathematics program in order to support students learning of the TEKS for mathematics.

Students. All children in participating classrooms were provided some form of CAI. However, only children whose parents provided active informed consent and whose families exclusively spoke English in the home were enrolled in the study. Participants' classmates who were learning English as a second language (i.e., 5% of peers) were allowed to participate in the testing if their parents provided consent, but their data were excluded from analyses because, as outliers, they would have undue influence on the findings. A total of 243 monolingual English-speaking children were enrolled: 127 female students and 120 male students. These children ranged in age at pretest from 5.04 to 6.71 years, (*M* = 5.62 years, *SD* = 0.32 years). At the study's onset, most participants achieved low average or below average scores on norm-referenced standardized tests of verbal ability (*M* = 84; *SD* = 14) and nonverbal ability (*M* = 77, *SD* = 11), indicating significant risk for poor academic outcomes. Most participating children represented ethnic minorities: 63% African American, 30% Hispanic, 4% mixed ethnicity, 2% Caucasian, and 1% other.

Research Design and Experimental Conditions

Study design. Participants were randomized with equal probability within each classroom to one of two conditions: CAI in mathematics delivered via the Building Blocks software or CAI in phonological awareness delivered via Earobics Step 1 (Version 1). Randomization at the student level to study conditions allowed us to identify the effect of the Building Blocks software on children's mathematics outcomes relative to those of children who participated in the Earobics Step 1 condition. It is important to note that none of the phonological awareness instruction delivered by Earobics Step 1 explicitly taught numeracy, quantity, geometric or spatial reasoning, or any other obvious mathematical skills. The study design ensured that children in the two experimental groups experienced the exact same classroom instruction and that group differences in academic outcomes were necessarily a consequence of the experimental conditions. Employing CAI in reading instruction as the control condition guaranteed that children in the control group did not receive any additional mathematics instruction during the time that children in the experimental group received CAI in mathematics. Also, the computerized nature of the control condition ensured that any positive gains associated with the Building Blocks software were not due to enhanced computer skills,

---

[1] The state curriculum standards are presented as part of the Texas Education Code (2012). The mathematics standards for Kindergarten can be found at http://ritter.tea.state.tx.us/rules/tac/chapter111/ch111a.pdf.

enhanced attentional abilities, increased motivation, or increased interactions with adults in a CAI context.

General CAI procedures. Children worked individually on computers in their school's computer lab during the ancillary instructional block designated for computer time. Although children worked individually with the computers, three research assistants were responsible for setting up hardware and software in each school's computer lab. Research assistants were also present in the lab to assist with noninstructional aspects of the supplemental CAI, such as providing behavioral supervision, technical assistance, and explanation of procedures if needed. For example, if a child had difficulty navigating between tasks, logging back into a program, or understanding a particular task, the research assistant provided appropriate direction. Research assistants were unaware of which children in the class were enrolled in the study.

All children used stereo headphones during CAI to help eliminate distraction and interference from background noise, given that CAI was delivered simultaneously to all children. All responses were made using an external mouse. Children in both groups received 90 minutes of CAI per week in addition to the standard instruction that they received in their general education classrooms. The 90 minutes of CAI per week were delivered in either three 30-minute sessions per week or two 45-minute sessions per week, depending on a given school's block schedule. The duration of 21 weeks of CAI was spread across 30 calendar weeks to accommodate the school districts' fall, winter, and spring holidays, district-wide standardized testing, and Kindergarten progress monitoring assessments.

It should be noted that the Building Blocks software and Earobics Step 1 are both adaptive software programs in that each adjusts the level of instruction to match the level of ability demonstrated by an individual child. In other words, each particular task (e.g., addition, sequencing, patterning, tapping sounds, blending sounds) is leveled such that a given task becomes more and more difficult until the child either successfully completes all levels of the task or chooses to discontinue the task by choosing to play a different game. Thus, children directed their own instruction insomuch as they were free to move from game to game within a given software program and by responding either correctly or incorrectly to each learning trial, which affected the level of instruction.

Computer software programs. The Building Blocks software teaches fundamental mathematical ideas through multiple series of leveled games that comprise a given learning trajectory along two separate strands, numeracy and geometry. Within each strand, there are a number of learning trajectories: counting, comparing and ordering numbers, subitizing, composing numbers, adding and subtracting, multiplying and dividing, classifying, measuring, recognizing shapes, composing shapes, comparing shapes, spatial sense and motions, and patterning. Participants in this study were not given access to any of the games that correspond to the geometry strand (e.g., recognizing shapes, comparing shapes, composing shapes) because of our focus on numeracy. Thus, participants were only permitted to interact with the games that teach numeracy, and all levels of those games were available (see the Appendix for the list of games and their description).

Earobics Step 1 includes six educational games that teach phonological awareness, short-term memory, sound discrimination, and letter–sound correspondence to children ages 4 to 7 years. Participants in this study were only permitted to interact with all levels of the three games that teach phonological awareness (i.e., Caterpillar Connection, Rhyme Time, and Rap-a-Tap-Tap).

In Caterpillar Connection, Katy-Pillar speaks parts of words (i.e., syllables, onset and rime, or phonemes), and children click on the picture that illustrates the blended word. If the correct picture is selected, then Katy turns into a butterfly, and praise and explanation are provided. Otherwise, Katy provides corrective feedback. Learning trials are leveled by increasing the number of response choices, the phonological similarity among response choices, and the length of time between parts of words that are to be blended together.

Rhyme Time includes two tasks. The rhyme-matching task has a character named Bog Frog who says a word, and children are instructed to select one of the other frogs that said a word that rhymed with Bog Frog's word. Children who master the rhyme-matching task proceed to the rhyme oddity task. For rhyme oddity, each frog speaks a word, and children are instructed to click on the frog whose word did not rhyme with the others. If the child responds incorrectly or does not respond within the time limit, then corrective feedback is provided, and the frog that represents the correct response choice jumps off its lily pad and into the water. When the child selects the correct frog, its lily pad sprouts a flower, and verbal praise and explanation are provided. For both rhyming tasks, teaching trials vary in number of response choices, phonological similarity among response choices, and loudness of background noise.

Rap-a-Tap-Tap teaches segmenting by having children click on an image of a drum each time a syllable or phoneme is spoken by an animated member of a rock band. If the child responds with the right number of clicks, then the rock band provides verbal praise and plays a short riff. If the child responds with the wrong number of clicks, a band member provides corrective feedback, and the drummer plays the correct number of taps. Instruction is leveled by increasing the number of word parts presented, decreasing the size of the linguistic units to attend to, and increasing the amount of time between the presentations of word parts.

Fidelity of implementation. Prior to the provision of any supplemental CAI, children's access privileges to a specific software suite and specific games were programmed according to their experimental conditions. Thereafter, a research assistant ensured that the CAI was implemented consistently and according to study design specifications by monitoring children's usage via daily reports generated online by each software program, which helped to ensure that children were moving along their learning trajectories within a software program. Research assistants checked all hardware, software, and power sources for proper functioning at the beginning of each day. They also supervised children's participation to ensure that children wore headphones, played on their own computer, and remained on task. Research assistants also maintained attendance records and logs of any technical problems experienced. Missed sessions were usually due to child absences, field trips, districtwide standardized testing, or technical difficulties such as Internet connectivity problems. Individual children or whole classes of children who missed a CAI session made up the missed session within a 2-week period.

Assessment Measures

Numeracy. Children's numeracy skills were assessed at the beginning and again at the end of the school year. Specifically, the fall administration (i.e., pretest) preceded CAI by 1 or 2 weeks, and the spring administration (i.e., posttest) followed CAI by 1 or 2 weeks. Numeracy skills were assessed with the Research-based Early Maths Assessment (REMA; Clements, Sarama, & Liu, 2008). The REMA was chosen because of its broad coverage of early numeracy skills and sensitivity to detect differences in early mathematics performance among young children (Clements et al., 2008). Items from the number concepts strand were administered because they indexed an outcome proximal to the intervention; however, items from the geometry strand were not administered because geometry was not addressed in this evaluation. The number strand includes four subscales: number recognition and subitizing, composition of number, arithmetic, and number comparison and sequencing. Core mathematics skills assessed within the number strand include verbal counting, object counting, number recognition and subitizing, number comparison, number sequencing, numeral recognition, number composition and decomposition, and adding and subtracting. General concepts and processes, such as part–whole thinking and the corresponding processes of composition and decomposition, classification, and seriation were woven throughout the core areas enumerated above. Standardized administration and scoring procedures were followed. Although the REMA is closely aligned conceptually with the Building Blocks software (i.e., proximal), it assesses numeracy skills more broadly and uses different tasks and materials so that it does not serve as an assessment specific to the Building Blocks software program. The internal consistency of this task with the present sample was good at pretest (alpha = .89) and posttest (alpha = .87).

Applied Problems. Because of our interest in examining treatment impact on distal mathematics outcomes, children were administered the Applied Problems subtest of the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2007). This subtest was administered at the end of the year during the first or second week that followed the completion of the intervention. The Applied Problems subtest requires children to analyze and solve verbally presented mathematics problems. For example, when looking at a stimulus page that depicts five ducks, two of which are swimming, the child is asked: "How many ducks are in the water?" The internal consistency of the Applied Problems subtest when used with the present sample at posttest was good (alpha = .90).

Vocabulary. To examine treatment specificity, that the effect of the Building Blocks Software was specific to mathematics and not vocabulary, children were administered the Expressive One-Word Picture Vocabulary Test (EOWPVT; Brownell, 2000). The EOWPVT presents examinees with colored line drawings that depict an action, object, category, or concept. Children were asked by an examiner to verbally respond to prompts such as "What is this?" "What is she doing?" and "What are these?" Standardized administration and scoring procedures were followed. The internal consistency of the EOWPVT when used with the present sample was good at pretest (alpha = .83) and posttest (alpha = .84).

Examiners. Seven experienced examiners administered the assessment battery. All examiners had undergraduate or advanced degrees. One was a former teacher, and another was a former psychologist. Examiners attended a 3-day workshop led by the second author. After training and ample practice, examiners demonstrated competence in the administration and scoring of all tests by administering the tests to the second author or to a postdoctoral fellow through role playing. Examiners were naïve to the study's aims and children's assignments to experimental conditions.

Results

Preanalysis Data Inspections

Of the 243 participants, 18 from the Building Blocks group and 15 from the Earobics group dropped out of the study (n = 33). Pretest comparisons indicated that these groups evidenced equivalent distributions for verbal ability ($F = 1.18$, $p = .28$), nonverbal ability ($F = 0.31$, $p = .58$), REMA number strand total scores ($F = 1.12$, $p = .29$), and three of the REMA subscale scores (F-values = 0.09 to 2.42, p-values = .12 to .77). Children who dropped out of the study, however, evidenced significantly higher scores on the arithmetic subscale of the REMA ($M = 1.38$, $SD = 3.57$) than children who remained in the study ($M = 0.53$, $SD = 1.27$) at pretest ($F = 5.79$, $p = .02$). Although these findings provide little evidence in support of group differences for key measures of verbal, nonverbal, and mathematics ability, listwise deletion excluded children who dropped out of the study in subsequent analyses.

The mathematics performance of children who completed the study is summarized by assessment wave for the full sample and by experimental condition in Table 1. Growth is apparent for the REMA number strand total score and most of its subscales. There was minimal evidence of data nonnormality. Floor effects were apparent for the arithmetic subscale at pretest for the Building Blocks group ($M = 0.39$, $SD = 1.04$, Skewness = 3.27) and the Earobics Step 1 group ($M = 0.85$, $SD = 1.36$, Skewness = 1.81). There was no additional evidence of nonnormality, and group variances were homogeneous.

To verify the success of random assignment, we evaluated whether or not the experimental groups differed on demographic characteristics and competencies at pretest. Tests of group differences indicated that the two experimental groups had equivalent distributions of age ($F = 0.74$, $p = .39$), gender (χ^2 (1) = 1.51, $p = .22$), and ethnicity[2] (χ^2 (2) = 1.65, $p = .44$). The two groups also had equivalent vocabularies ($F = 1.88$, $p = .17$) and nonverbal abilities ($F = 1.31$, $p = .25$). For mathematics, the two groups evidenced equivalent pretest distributions for the REMA number strand total score ($F = 1.55$, $p = .21$). Of the REMA subscales, the groups differed for arithmetic ($F = 4.02$, $p = .05$) in favor of the Earobics Step 1 group (see Table 1 for achievement scores). The groups did not differ significantly on the three remaining subscales of the REMA (F-values = 0.59 to 3.62, p-values = .06 to .44) at pretest.

Finally, unconditional means models and means models conditional on pretest mathematics achievement were estimated for the full sample to compute intraclass correlations (ICC) and design effects (DE) for posttest scores. The ICCs (see Table 2) for the unconditional means models ranged from .00 to .08, indicating that between 0% and 8% of the variance in mathematics scores at posttest were due to classroom-level differences. DEs ranged from 1.00 to 4.19, one of which exceeded a recommended cutoff criterion of 2.0 (cf. Muthén, 1991, 1994), suggesting the need to use multilevel modeling. For the conditional means models, ICCs ranged from .00 to .15. Corresponding DEs ranged from 1.00 to 6.62, two of which exceeded the cutoff criterion of 2.0. Given the ICCs, DEs, and desire to maintain consistency across analyses, multilevel modeling was employed when analyzing all outcomes to guard against Type 1 error and biased parameter estimates (cf. Peugh, 2010; Singer & Willett, 2003).

General Data Analytic Approach

Multilevel modeling afforded us the opportunity to consider the effect of experimental condition while accounting for classroom-level differences (e.g., variations in teacher styles, instructional emphases) on assessment results. Within this context, the influence of experimental condition on mathematics outcomes was evaluated separately because a univariate approach was consistent with our interest in examining treatment specificity (e.g., impact on numeracy but not vocabulary) versus impacts on proximal (REMA) and distal (Applied Problems) measures. The direct effect of wave on REMA scores was tested using repeated measures (pre- and post-intervention). The intervention's impact on both REMA scores and Applied Problems scores was then investigated with a multilevel ANCOVA. Specifically, pretest scores were specified as the covariate and experimental condition was specified as the predictor of children's mathematics outcomes. All models were estimated with restricted maximum likelihood and the Kenward-Roger method for tests of fixed effects.[3] Finally, multilevel effect

---

[2] Because of the limited frequency of "mixed ethnicity," "Caucasian," and "other," these three categories were combined to form a single group for the purpose of evaluating the distribution of ethnicity between treatment groups.

Table 1

Mathematics Means and Standard Deviations for the Full Sample and by Experimental Condition by Assessment Wave

Measure	Max	Fall		Spring	
		M	*SD*	*M*	*SD*
Full sample					
Number strand total	43	13.40	6.01	20.36	6.67
Number recognition and subitizing	5	2.64	1.06	3.36	0.93
Composition of number	5	1.06	1.41	1.76	1.83
Arithmetic	15	0.63	1.23	2.01	2.37
Number comparison and sequencing	18	9.72	4.11	13.77	3.59
Applied Problems	63	—[a]	—[a]	18.45	3.70
Building Blocks					
Number strand total	43	12.91	5.43	20.79	6.73
Number recognition and subitizing	5	2.50	1.11	3.47	0.91
Composition of number	5	0.94	1.43	2.07	1.98
Arithmetic	15	0.39	1.04	2.23	2.47
Number comparison and sequencing	18	9.52	3.73	13.65	3.45
Applied Problems	63	—[a]	—[a]	18.76	3.66
Earobics Step 1					
Number strand total	43	13.89	6.52	19.91	6.62
Number recognition and subitizing	5	2.77	1.00	3.25	0.94
Composition of number	5	1.18	1.40	1.47	1.62
Arithmetic	15	0.85	1.36	1.78	2.25
Number comparison and sequencing	18	9.92	6.52	13.89	3.73
Applied Problems	63	—[a]	—[a]	18.13	3.73

Note. All values are reported in raw score units without corrections for pretest differences or classroom nesting.

[a] The Applied Problems measure was not administered in the fall of the school year.

Table 2
Intraclass Correlations and Design Effects

	Unconditional means models		Conditional means models[a]	
	ICC	DE	ICC	DE
Number strand total	.000	1.00	.013	1.49
Number recognition and subitizing	.001	1.04	.010	1.38
Composition of number	.019	1.72	.000	1.00
Arithmetic	.000	1.00	.000	1.00
Number comparison and sequencing	.000	1.00	.066	3.51
Applied Problems[b]	.084	4.19	.148	6.62

Note. ICC is intraclass correlation; DE is design effect.

[a] Conditional on pretest achievement.

[b] Number strand total score was used as pretest achievement because the Applied Problems measure was not administered in the fall of the school year.

size estimates were determined using calculation procedures derived from Feingold (2009),[4] which are consistent with Raudenbush and Liu (2001).

Mathematics Learning and Building Blocks Software Impact

REMA number strand total score. A repeated measures examination of the effect of wave for the full sample indicated that number strand total scores at post-test were significantly higher than at pretest after accounting for classroom nesting, F(1, 412) = 141.77, $p < .0001$. The average improvement for the sample as a whole was 6.94 points on this 43-item scale. Next, a multilevel ANCOVA that included REMA number strand total scores at pretest and group found that experimental condition significantly predicted number strand total scores at post-test, F(1, 178) = 8.08, $p < .01$, after controlling for individual differences in pretest numeracy. The Building Blocks group outperformed the comparison group, with a difference in least squared means of 1.85 raw score units. The resulting effect size was 0.43, which exceeds the WWC threshold of a substantively important positive effect (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2014).

---

[3] Degrees of freedom based on the Kenward-Roger method.

[4] The following formula adapted from Feingold (2009) was used to determine estimates of effect size: $ES = \beta_{11}\,(time)/SD_{RAW}$, where β_{11} (average growth rate) is the treatment effect accounting for the multilevel structure of the data, SD_{RAW} is the treatment effect's standard deviation, and *time* is the number of time points or waves of data. This method conveys effect magnitude by estimating the difference between the treatment groups' mean growth rates and is calculated with the standard deviation of raw scores.

REMA number recognition and subitizing. Again, a repeated measures examination of the effect of wave for the full sample indicated that number recognition and subitizing scores at posttest were significantly higher than at pretest after accounting for classroom nesting, $F(1, 423) = 59.70$, $p < .0001$. The average improvement for the sample as a whole was 0.71 points on this 5-item subscale. Next, a multilevel ANCOVA that included pretest scores on this subscale and group as predictors indicated that experimental condition significantly predicted posttest number recognition and subitizing scores. As with the number strand total scores, the results indicated that the Building Blocks group outperformed the comparison group, $F(1, 188) = 6.07$, $p = .01$, with a difference in least squared means of 0.30 raw score units. The resulting effect size was 0.36, which represents a substantively important positive effect.

REMA composition of number. The effect of wave on the composition of number subscale for the full sample was significant after accounting for classroom nesting, $F(1, 348) = 18.08$, $p < .0001$, indicating that scores for composition of number were significantly higher at posttest than at pretest. The average improvement for the sample as a whole was 0.72 points on this 5-item subscale. Next, a multilevel ANCOVA indicated that experimental condition significantly predicted composition of number posttest scores when accounting for pretest scores on this subscale, $F(1, 134) = 7.61$, $p < .01$. The Building Blocks group outperformed the comparison group, with a difference in least squared means of 0.82 raw score units. The resulting effect size was substantively important, 0.48.

REMA arithmetic. The effect of wave on the arithmetic subscale was significant after accounting for classroom nesting, $F(1, 277) = 34.70$, $p < .0001$. The average improvement for the sample as a whole was 1.38 points on this 15-item arithmetic subscale. Next, a multilevel ANCOVA that accounted for pretest arithmetic scores indicated that experimental condition significantly predicted arithmetic posttest scores, $F(1, 96) = 7.77$, $p < .01$. Again, the Building Blocks group outperformed their comparison peers, with a difference in least squared means of 1.36 raw score units. The resulting effect size was substantively important, 0.57.

REMA number comparison and sequencing. After accounting for classroom nesting, the effect of wave was significant, $F(1, 417) = 127.31$, $p < .0001$. The average improvement for the sample as a whole was 4.03 points on the 18-item number comparison and sequencing subscale. Results of the next analysis, contrary to the results for other mathematics outcomes, indicated that the experimental condition did not explain variation in posttest number comparison and sequencing scores beyond that explained by pretest scores on this subscale, $F(1, 181) = 0.01$, $p = .94$. The Building Blocks group performed similarly to the comparison group, with a difference in least squared means of 0.03 raw score units. The resulting effect size was 0.01. Thus, all students improved in number comparison and sequencing between pretest and posttest; however, there was not a reliable difference between groups.

Woodcock-Johnson Applied Problems. The effect of wave on change in scores obtained on the Applied Problems test was not examined because this measure was only administered at posttest. When the number strand total score was specified as a pretest covariate in the multilevel ANCOVA that evaluated treatment impact, results indicated a statistically significant effect for the experimental condition after accounting for pretest number strand total scores and classroom nesting, $F(1, 176) = 5.90$, $p = .02$. Results favored the Building Blocks group over the control group, with a difference in least squared means of 0.96 raw score units. This equates to a substantively important effect size of 0.37 and a difference in least squared means of the norm-referenced standard scores of approximately 3.61 standard score units.

Vocabulary Learning and Building Blocks Software Impact

To examine treatment specificity, we examined the effect of the Building Blocks software on children's vocabulary scores. A repeated measures examination of the effect of wave for the full sample indicated that vocabulary scores at posttest were significantly higher than at pretest after accounting for classroom nesting, $F(1, 422) = 78.81$, $p < .001$. Average improvement for the sample as a whole was 9.49 points on this 130-item test. However, results of the subsequent multilevel ANCOVA indicated that the experimental condition did not explain variation in posttest vocabulary scores beyond that explained by pretest scores on this test when controlling for classroom nesting, $F(1, 198) = 0.25$, $p = .62$. The Building Blocks group performed similarly to the comparison group, with a difference in least squared means of 2.06 raw score units. The resulting effect size was -0.07. Thus, all students improved in vocabulary between pretest and posttest; however, group differences were not reliable. It can therefore be concluded that the Building Blocks software demonstrated treatment specificity, impacting mathematics skills but not vocabulary.

Discussion

The purpose of this study was to examine the effect of adding the Building Blocks software to general education mathematics instruction in schools that serve children from low-income and ethnic minority backgrounds. Our sample of children had low average or below-average scores on standardized tests of verbal and nonverbal ability when they entered Kindergarten, confirming their at-risk status. The findings demonstrated that a relatively low-intensity supplemental implementation of the Building Blocks software throughout most of the Kindergarten school year led to reliable improvements in mathematics achievement. Specifically, the experimental group who received supplemental CAI via the Building Blocks software performed significantly higher on posttest measures of numeracy and Applied Problems than did the control group who received supplemental CAI via Earobics Step 1, after controlling for pretest numeracy achievement. The value added by the Building Blocks software to children's numeracy and applied mathematics skills was moderate (effect size = 0.43 and 0.37, respectively) and exceeded the WWC threshold of a 0.25 effect size to be considered of "substantive importance" (U.S. Department of Education, Institute of Education Sciences, What

Works Clearinghouse, 2014). The present findings are also commensurate with results from research in CAI (Fletcher-Flinn & Gravatt, 1995; Lavin & Sanders, 1983; Niemiec & Walberg, 1985; Ragosta et al., 1982), more recent research studies (Clements & Sarama, 2003, 2010; Räsänen et al., 2009), and meta-analyses (Cheung & Slavin, 2013; Li & Ma, 2010; Slavin & Lake, 2008), which collectively suggest that the mathematics learning of Kindergarten children can be enhanced by using research-based software interventions. Nonetheless, non-CAI can produce similar, if not stronger effects in samples of children similar to those in the present study. Dyson, Jordan, Beliakoff, and Hassinger-Das (2015) demonstrated that a supplemental Kindergarten number sense intervention delivered to small groups of students had substantially positive impacts on children's number sense, arithmetic fluency, and mathematics calculation achievement (effect sizes = 0.32 to 0.82). In contrast, Clarke et al. (2015) reported negligible effects (p-values = .0517 to .89) of a core Kindergarten mathematics curriculum on four mathematics outcomes (effect sizes = -0.008 to .108). The effect sizes in the present study are therefore significant because supplemental use of the Building Blocks software, as implemented in the present study, does not involve a large amount of teachers' time, does not interrupt children's regular mathematics classroom instruction, and does not interfere with other academic programming.

A noteworthy pattern of findings was that the Building Blocks software was generally found to have its greatest effects on mathematics outcomes measured by the REMA. This was not surprising, given that the REMA was designed to assess all of the progressions in mathematics development that underlie the Building Blocks learning trajectories (see Sarama & Clements, 2009). Although the REMA is closely aligned conceptually with the Building Blocks software, it was developed before the software and assesses numeracy skills using different tasks and different materials; thus, it does not exclusively serve as a curriculum mastery test. Nonetheless, we included the Applied Problems subtest as a more distal measure to assess generalization of learned skills to broader mathematics achievement. This subtest is not aligned with the Building Blocks software, and it assesses a broader skill set as a general outcome measure. A possible alternative explanation of the pattern of results is that perhaps the REMA is more sensitive than the Applied Problems subtest at measuring Kindergarteners' mathematics abilities. For instance, the descriptive statistics of children's raw scores on the two measures at postintervention indicate greater variance in scores on the REMA, which could be (but are not necessarily) indicative of increased sensitivity to individual differences. Indeed, two national panels on preschool assessment (National Institute of Child Health and Human Development, 2002; Kochanoff, Hirsh-Pasek, Newcombe, & Weinraub, 2003) cautioned against the sole use of the Woodcock-Johnson test for assessment of mathematical skills in preschoolers (and Kindergartners) because it has not been validated for children in the youngest age ranges; covers a narrow range of problems (e.g., the oft-used Applied Problems has multiple tasks in which children must count a proper subset, all with numbers from 1 to 4) and jumps too quickly to advanced, formal knowledge; and is not based on current research on the development of mathematical thinking, including giving little attention to developmental sequences. Regardless of which possible explanation of the noted pattern of findings holds true, the fact that there were reliable group differences on Applied Problems at postintervention only strengthens the conclusion that the Building Blocks software improved mathematics achievement because either the intervention effects generalized to broad application of mathematics skills or the intervention effects were robust enough to be evident on a less sensitive outcome measure.

When considering the effects of the Building Blocks software on specific mathematics skills, the software program was shown to positively affect children's skills with arithmetic, number composition, and number recognition and subitizing. With the exception of the null effect on number comparison and sequencing, the positive effects of the Building Blocks software evidenced in the present study are consistent with findings from previous research that evaluated the larger, comprehensive Building Blocks program (e.g., Clements & Sarama, 2007, 2008; Sarama & Clements, 2009). In considering the one null finding, it is important to note that the teacher-led instruction that takes place as part of the larger Building Blocks program includes an emphasis on higher level thinking. Moreover, teacher-led instruction with the full Building Blocks program generally targets number comparison and sequencing more frequently and intensely than does the software (see Clements et al., 2011). Thus, the active involvement of the teacher, whether through whole-group or small-group instruction, is a critical component for engaging children in mathematical reasoning that should not be replaced by CAI.

Finally, although it is important not to overgeneralize, it is worth noting that the software that we evaluated was designed on scientific principles. That is, the Building Blocks software was based on a refined version of a previously published model for "developing effective software" (Clements & Battista, 2000). The final model is a 10-phase research-and-development process with three categories of development, a priori research reviews, the development of learning trajectories, and formative and summative evaluation (described in Clements, 2007, and Clements & Sarama, 2015). However, all summative evaluations before the present study were of the entire set of Building Blocks materials. Results concerned with the effectiveness of the software suite were therefore confounded with the effectiveness of the larger Building Blocks program. This is the first study to validate this specific software suite and thus provides evidence of the efficacy of the scientific principles for software development.

Implications

Children at risk of school failure need high-quality instructional experiences to prevent or mitigate the development of mathematics learning difficulties (Bowman, Donovan, & Burns, 2001; Cross et al., 2009; Magnuson et al., 2004; Peisner-Feinberg et al., 2001; Shonkoff & Phillips, 2000). The present study provides support for the use of the Building Blocks software as a supplemental mathematics program in an effort to decrease risk of school failure by increasing the mathematical competencies of children from low-income and ethnic minority backgrounds. Further, the software program's value-added and broad impacts speak well of the program's adaptive functionality and comprehensive coverage of relevant competencies. That is, even though children seemingly directed their own instruction, the adaptive algorithms assured that instruction was provided at appropriate levels such that it allowed children to extract benefits from the educational software in accord with their personal competencies. Breadth of scope and adaptive instruction are important considerations for both developers and consumers of educational software. Administrators, educators, and parents alike desire educational software that teaches multiple competencies and that can appropriately be used with children who vary in school readiness. Thus, use of the Building Blocks software can provide substantial benefits regarding mathematics learning, at least when used as a supplement to general education.

A specific policy implication that follows from the present results is that CAI interventions may be especially useful in diverse schools, such as those in this study. Given the effectiveness of the Building Blocks software and its ease of implementation, which does not interfere with children's academic programming, further use of the software in the prevention of mathematics learning difficulties is supported. In particular, schools and districts are continuously faced with how best to meet the educational needs of their students in the context of limited resources. Such software offers much promise as an efficient, affordable, and effective form of intervention and prevention because it precisely provides instruction at the level of a given pupil's ability, and it does so without an unfeasible 1:1 ratio of educators to students. Thus, these results contradict concerns about the use of CAI with students from low-income and minority backgrounds (Kitchen & Berk, 2016). These cautionary polemics are examples of what Papert called *technocentrism* (Papert, 1987), which refers to the focus on the technology rather than focusing on the pedagogical context as a whole, including the content of particular CAI programs. Such perspectives can oversimplify challenges faced within educational systems, such as access to high-quality and rigorous mathematics education in the United States, by limiting discussions of CAI to programs focused on drill and skill development and then generalizing to other forms of educational software (Kitchen & Berk, 2016). The Building Blocks software studied here includes different goal and pedagogical structures, which may account for its success with this low-income population. That is, its design avoids the false and pernicious division between skills and reasoning, which we believe is more consistent with the CCSSM than is reflected in many polemics. Although the present findings are valuable to administrators who allocate intervention resources, the Building Blocks software should not be used as a band-aid or "fix all" to compensate for classroom instruction that lacks intensity. Instead, it should be implemented judiciously as a supplement, and its effectiveness under various instructional contexts should be studied further.

Limitations

First, the present findings may not be generalized beyond the population of ethnic minority Kindergarten children in low-income schools in Texas. For example, the impact of the Building Blocks software may differ in other populations of children, such as those who evidence developmental delays, learning disabilities, or intellectual disabilities; children identified as speech, language, or hearing impaired; or children learning English as a second language. Second, the present study only evaluated the efficacy of the Building Blocks software games that teach early numeracy. The Building Blocks software games that teach geometric understandings, such as identifying shapes and their components, composition of their shape, representing shape, geometric measurement, patterning, and comparing shapes, were neglected in this randomized control trial. Third, although the study design permits causal statements about the impact of supplemental use of the Building Blocks software, the design does not permit statements concerning the merit of the Building Blocks software relative to other supplemental mathematics programs. For example, it remains possible that other supplemental mathematics programs, tutoring, and even mathematics homework could be equally beneficial. Thus, improved mathematics outcomes of children in the Building Blocks software condition may in part be a result of additional time practicing mathematics. Finally, it is unclear how regular classroom instruction supported or impeded the learning of mathematics of the children in the present study. For example, some regular classroom teachers may have been more effective in teaching mathematics relative to other teachers. Although we accounted for such classroom effects in our research design and statistical models, the present study was not designed to examine the interaction of supplemental use of the Building Blocks software and general education programming. Instead, the generalization of the study's findings is limited to use of the Building Blocks software as a supplement to the enVisionMath curriculum.

Conclusion

In summary, the Building Blocks software led to improved numeracy and applied problems achievement in monolingual English-speaking ethnic minority children from low-income backgrounds. These results are consistent with prior research and add to evaluations that support the judicious use of research-based educational software. Similarly, these findings suggest that adaptive mathematics instruction software programs may be an efficacious supplemental method for ameliorating mathematics learning difficulties in children from ethnic minority and low-income backgrounds.

References

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher, 36*(7), 369–387. doi:10.3102/0013189X07308739

Anthony, J. L. (2015). For which children of economic disadvantage and in which instructional contexts does Earobics Step 1 improve kindergarteners' literacy? *Journal of Research on Educational Effectiveness*. Advance online publication. doi:10.1080/19345747.2015.1055637

Baroody, A. J., Eiland, M. D., Pupura, D. J., & Reid, E. E. (2013). Can computer-assisted discovery learning foster first graders' fluency with the most basic addition combinations? *American Educational Research Journal, 50*(3), 533–573. doi:10.3102/0002831212473349

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*(2), 115–130. doi:10.1086/525550

Bowman, B. T., Donovan, M. S., & Burns, M. S. (Eds.). (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press. doi:10.17226/9745

Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report, 17*(1). Retrieved from http://srcd.org/sites/default/files/documents/spr17-1.pdf

Brownell, R. (2000). *Expressive One-Word Picture Vocabulary Test manual*. Novato, CA: Academic Therapy Publications.

Building Blocks [Computer software]. Columbus, OH: SRA/McGraw-Hill.

Case, R., Griffin, S., & Kelly, W. M. (1999). Socioeconomic gradients in mathematical ability and their responsiveness to intervention during early childhood. In D. P. Keating & C. Hertzman (Eds.), *Developmental health and the wealth of nations: Social, biological, and educational dynamics* (pp. 125–149). New York, NY: Guilford Press.

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K–12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88–113. doi:10.1016/j.edurev.2013.01.001

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review, 28*(4), 415–427. doi:10.1016/j.econedurev.2008.09.003

Clarke, B., Baker, S., Smolkowski, K., Doabler, C., Strand Cary, M., & Fien, H. (2015). Investigating the efficacy of a core kindergarten mathematics curriculum to improve student mathematics learning outcomes. Journal of Research on Educational Effectiveness, 8(3), 303–324. doi:10.1080/19345747.2014.980021

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education, 38*(1), 35–70. doi:10.2307/30034927

Clements, D. H., Baroody, A. J., & Sarama, J. (2013). *Background research on early mathematics.* Washington, DC: National Governors Association. Retrieved from http://www.nga.org/files/live/sites/NGA/files/pdf/2013/1311SEME-Background.pdf

Clements, D. H., & Battista, M. T. (2000). Designing effective software. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 761–776). Mahwah, NJ: Erlbaum.

Clements, D. H., & Sarama, J. (2003). Strip mining for gold: Research and policy in educational technology—A response to "Fool's Gold." *Educational Technology Review, 11*(1), 7–69.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the *Building Blocks* project. *Journal for Research in Mathematics Education, 38*(2), 136–163. doi:10.2307/30034954

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*(2), 443–494. doi:10.3102/0002831207312908

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

Clements, D. H., & Sarama, J. (2010). Technology. In V. Washington & J. D. Andrews (Eds.), *Children of 2020: Creating a better tomorrow* (pp. 119–123). Washington, DC: Council for Professional Recognition/National Association for the Education of Young Children.

Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. *Science, 333*(6045), 968–970. doi:10.1126/science.1204537

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

Clements, D. H., & Sarama, J. (2015). Methods for developing scientific education: Research-based development of practices, pedagogies, programs, and policies. In O. N. Saracho (Ed.), *Handbook of research methods in early childhood education: Review of research methodologies* (Vol. 1, pp. 717–751). Charlotte, NC: Information Age.

Clements, D. H., & Sarama, J. (in press). Promoting a good start: Technology in early childhood mathematics. In E. Arias, J. Cristia, & S. Cueto (Eds.), *Promising models to improve primary mathematics learning in Latin America and the Caribbean using technology*. Washington, DC: Inter-American Development Bank.

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*(4), 457–482. doi:10.1080/01443410701777272

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education, 42*(2), 127–166. doi:10.5951/jresematheduc.42.2.0127

Cross, C. T., Woods, T. A., & Schweingruber, H. (Eds.). (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. Washington, DC: National Academies Press. doi:10.17226/12519

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1880–1990* (2nd ed.). New York, NY: Teachers College Press.

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

Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade* (NCES 2002–125). Washington, DC: National Center for Education Statistics.

Dyson, N., Jordan, N. C., Beliakoff, A., & Hassinger-Das, B. (2015). A kindergarten number-sense intervention with contrasting practice conditions for low-achieving children. *Journal for Research in Mathematics Education, 46*(3), 331–370. doi:10.5951/jresematheduc.46.3.0331

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., . . . Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort* (NCEE 2007-4005). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Earobics Step 1 (Version 1) [Computer software]. Boston, MA: Houghton Mifflin Harcourt.

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods, 14*(1), 43–53. doi:10.1037/a0014699

Fletcher-Flinn, C. M., & Gravatt, B. (1995). The efficacy of computer assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research, 12*(3), 219–241. doi:10.2190/51D4-F6L3-JQHU-9M31

Friedman-Krauss, A., & Barnett, S. W. (2013, April). *Early childhood education: Pathways to better health* (NIEER Policy Brief, Issue 25). New Brunswick, NJ: National Institute for Early Education Research.

Ginsburg, A., Leinwand, S., Anstrom, T., & Pollock, E. (2005). *What the United States can learn from Singapore's world-class mathematics system (and what Singapore can learn from the United States): An exploratory study*. Washington, DC: American Institute for Research.

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

Harskamp, E. (2015). The effects of computer technology on primary school students' mathematics achievement: A meta-analysis. In S. Chinn (Ed.), *The Routledge international handbook of dyscalculia and mathematical learning difficulties* (pp. 383–392). New York, NY: Routledge.

Horn, C. (2005). Standardized assessments and the flow of students into the college admission pool. *Educational Policy, 19*(2), 331–348. doi:10.1177/0895904804274057

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*(3), 850–867. doi:10.1037/a0014939

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press. doi:10.17226/9822

Kitchen, R. S., & Berk, S. (2016). Educational technology: An equity challenge to the Common

Core. *Journal for Research in Mathematics Education, 47*(1), 3–16. doi:10.5951/jresematheduc.47.1.0003

Kochanoff, A. T., Hirsh-Pasek, K., Newcombe, N., & Weinraub, M. (2003). *Using science to inform preschool assessment: A summary report of the Temple University Forum on Preschool Assessment*. Philadelphia, PA: Temple University, Center for Improving Resources in Children's Lives.

Krajewski, K. (2005, April). *Prediction of mathematical (dis-)abilities in primary school: A 4-year German longitudinal study from kindergarten to grade 4*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Atlanta, GA.

Lavin, R. J., & Sanders, J. E. (1983). Longitudinal evaluation of the C/A/I computer assisted instruction Title 1 project: 1979–82. Chelmsford, MA: Merrimack Education Center.

Lentz, C. L., Seo, K. K.-J., & Gruner, B. (2014). Revisiting the early use of technology: A critical shift from "how young is too young?" to "how much is 'just right'?" *Dimensions of Early Childhood, 42*(1), 15–23.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22*(3), 215–243. doi:10.1007/s10648-010-9125-8

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal, 41*(1), 115–157. doi:10.3102/00028312041001115

Moradmand, N., Datta, A., & Oakley, G. (2013). My maths story: An application of a computer-assisted framework for teaching mathematics in the lower primary years. In R. McBride & M. Searson (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2013* (pp. 3299–3307). Chesapeake, VA: Association for the Advancement of Computing in Education.

Moyer-Packenham, P. S., & Westenskow, A. (2013). Effects of virtual manipulatives on student achievement and mathematics learning. *International Journal of Virtual and Personal Learning Environments, 4*(3), 35–50. doi:10.4018/jvple.2013070103

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28*(4), 338–354. doi:10.1111/j.1745-3984.1991.tb00363.x

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22*(3), 376–398. doi:10.1177/0049124194022003006

National Assessment of Educational Progress [NAEP]. (2013). The nation's report card: Which student groups are making gains? Retrieved from http://www.nationsreportcard.gov/reading_math_2013/#/gains-by-group

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Author. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

National Institute of Child Health and Human Development. (2002). *Early childhood education and school readiness: Conceptual models, constructs, and measures*. Retrieved from https://www.nichd.nih.gov/publications/pubs/documents/school_readiness.pdf

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

Niemiec, R. P., & Walberg, H. J. (1985). Computers and achievement in the elementary schools. *Journal of Educational Computing Research, 1*(4), 435–440. doi:10.2190/7TP6-GYVE-1V8F-RGV7

Nusir, S., Alsmadi, I., Al-Kabi, M., & Sharadgah, F. (2013). Studying the impact of using multimedia interactive programs on children's ability to learn basic math skills. *E-Learning and Digital Media, 10*(3), 305–319. doi:10.2304/elea.2013.10.3.305

Papert, S. (1987). Computer criticism vs. technocentric thinking, *Educational Researcher, 16*(1), 22–30. doi:10.3102/0013189X016001022

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*(5), 1534–1553. doi:10.1111/1467-8624.00364

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85–112. doi:10.1016/j.jsp.2009.09.002

Ragosta, M., Holland, P. W., & Jamison, D. T. (1982). *Computer-assisted instruction and compensatory education: The ETS/LAUSD study. The final report*. Princeton, NJ: Educational Testing Service.

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development, 24*(4), 450–472. doi:10.1016/j.cogdev.2009.09.003

Raudenbush, S. W., & Liu, X.-F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6*(4), 387–401. doi:10.1037/1082-989X.6.4.387

Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J., . . . Martinez, M. E. (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-temporal math. *Journal of Research on Educational Effectiveness, 7*(4), 358–383. doi:10.1080/19345747.2013.856978

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

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

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195152968.001.0001

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427–515. doi:10.3102/0034654308317473

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly, 19*(1), 99–120. doi:10.1016/j.ecresq.2004.01.002

Texas Education Code, §§7.102(c)(4), 28.002, and 28.0021(a)(1) (2012). Retrieved from http://ritter.tea.state.tx.us/rules/tac/chapter111/ch111a.pdf

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013). *Elementary school mathematics intervention report: DreamBox Learning*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_dreambox_121013.pdf

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2014). *What Works Clearinghouse: Procedures and standards handbook* (Version 3.0). Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (n.d.). *Find What Works*. Retrieved from http://ies.ed.gov/ncee/wwc/findwhatworks.aspx

Woodcock, R. W., McGrew, K. S., & Mather, N. (2007). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing Company.

Authors

Matthew E. Foster, Children's Learning Institute, University of Texas Health Science Center, 7000 Fannin Street, Suite 2373D, Houston, TX 77030; Matthew.E.Foster@uth.tmc.edu

Jason L. Anthony, Children's Learning Institute, University of Texas Health Science Center at Houston, 7000 Fannin Street, Suite 2377, Houston, TX 77030; Jason.L.Anthony@uth.tmc.edu

Doug H. Clements, University of Denver, Katherine R. Ruffatto Hall 154, 1999 East Evans Avenue, Denver, CO 80208-1700; Douglas.Clements@du.edu

Julie Sarama, University of Denver, Katherine R. Ruffatto Hall 154, 1999 East Evans Avenue, Denver, CO 80208-1700; Julie.Sarama@du.edu

Jeffrey M. Williams, Children's Learning Institute, University of Texas Health Science Center at Houston, 7000 Fannin Street, Suite 2373, Houston, TX 77030; Jeffrey.Williams@uth.tmc.edu

Submitted April 16, 2015
Accepted August 26, 2015

APPENDIX

Table A1
Description of Building Blocks Software Games Targeting Numeracy

Game	Mathematics skills targeted	Description
Number Snapshots	Number recognition and subitizing (instantly recognize)	Children are given only a few seconds to see a digital image (an array of dots) before they have to click on the target numeral on the film strip to the left (or click on the peek button for another look) of the image.
Before and After Math	Verbal counting and arithmetic	Children identify and select numbers that come either just before (i.e., $n - 1$) or right after (i.e., $n + 1$) a target number.
Sea to Shore	Verbal counting and arithmetic	Children identify number amounts by moving forward (or counting on) a number of spaces on a game board (image displayed on the computer screen) that is *one more* than a given numeral.
Book Stacks	Verbal counting and arithmetic	Children "count on" (through at least one decade) from a given number as they load books onto a cart.
Memory Number	Number recognition, composition of number, and number comparison	Children match displays containing numerals with displays containing collections of objects within the framework of a "Concentration" card game.
Number Compare	Number comparison and sequencing	Children compare two cards and then choose the one with the greater value.
School Supply Shop	Number sequencing and skip counting	Children see it as counting school supplies that are bundled in groups of ten to reach a target number.
Tire Recycle	Number sequencing and skip counting	Children count tires by twos or fives as they move the groups of tires into a recycling container.
Bright Idea	Verbal counting, counting strategies, and arithmetic	Children count on from a numeral to identify number amounts and then move forward a corresponding number of spaces on a game board.