Early Childhood Research Quarterly 43 (2018) 83–95

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Hispanic dual language learning kindergarten students' response to a numeracy intervention: A randomized control trial

Matthew E. Foster[a,b,*], Jason L. Anthony[a,b], Doug H. Clements[c], Julie Sarama[c], Jeffrey J. Williams[a,b]

[a] Rightpath Research and Innovation Center, University of South Florida, 13301 Bruce B. Downs Blvd., MHC 1706, Tampa, FL 33602, United States
[b] University of Denver, Katherine R. Ruffatto Hall 154, 1999 East Evans Avenue, Denver, CO 80208-1700, United States
[c] Rightpath Research and Innovation Center, University of South Florida, 13301 Bruce B. Downs Blvd., MHC 1717, Tampa, FL 33602, United States

ARTICLE INFO

Article history:
Received 28 March 2017
Received in revised form 16 January 2018
Accepted 28 January 2018

Keywords:
Dual language learners
Kindergarten mathematics
Computer-assisted instruction
Numeracy
Hispanic students

ABSTRACT

This study evaluated the impact of the Spanish version of the Building Blocks software program and vocabulary on kindergarten mathematics outcomes. Participants included 270 Hispanic dual language learners from low-income communities. Relative to children in the computer assisted instruction (CAI) literacy control group, those in the Building Blocks CAI group evidenced higher posttest scores for Spanish mathematics, but not for English mathematics, after controlling for pretest numeracy. There were also main effects of English vocabulary and Spanish vocabulary predicting posttest mathematics scores in English and Spanish, after controlling for covariates. These findings support the use of the Building Blocks software as a supplemental method for improving the mathematics competencies of Hispanic dual language learners from low-income backgrounds.

1. Introduction

Relative to non-Hispanic White peers, Hispanic students in the U.S. have underperformed in mathematics for decades (Clements & Sarama, 2011; Denton & West, 2002; National Center for Educational Statistics, 2015; Natriello, McDill, & Pallas, 1990). Hispanic students are at increased risk for low academic achievement due to high rates of poverty (Lopez & Velasco, 2011) and lower rates of preschool enrollment (Barnett, Carolan, & Fitzgerald, & Squires, 2011; Figueras-Daniel & Barnett, 2013; Kena et al., 2016) than peers from similar economic backgrounds. Some also argue that low levels of English language proficiency among Hispanic children learning Spanish and English (i.e., Hispanic dual language learners) is at least partially responsible for these children's low rates of 'academic success' in the U.S. (Galindo, 2010). Given the experiential gaps to acquire English proficiency and that 'academic success' in the U.S. will probably continue to be defined as achievement on standardized tests administered in English despite the growing embracement of dual language instruction in some states, there is a pressing need to identify mathematics programs that promote mathematics competencies in Spanish and English among Hispanic dual language learners (DLLs). Therefore, the present study examines the impact of a numeracy intervention when used with Hispanic DLLs and considers the effect of vocabulary on kindergarten mathematics outcomes.

2. Relations between language and mathematics

Language is the medium of classroom instruction and is thought by some to be the means by which young children refine their understanding of numbers (Purpura, Napoli, Wehrspann, & Gold, 2016; Spelke, 2003). Indeed, empirical study has demonstrated that scores of English language at school entry are predictive of growth in scores on English mathematics tests through ninth grade (Duncan et al., 2007; Hooper, Roberts, Sideris, Burchinal, & Zeisel, 2010; Purpura, Hume, Sims, & Lonigan, 2011; Romano, Babchishin, Pagani, & Kohen, 2010). However, few studies that consider the prediction of kindergarten mathematics outcomes from language scores at school entry include Hispanic DLLs. For example, Kleemans, Segers, and Verhoeven (2011) demonstrated that second

---

The research reported in this manuscript and initial drafts of this manuscript were completed at the Children's Learning Institute at the University of Texas Health Science Center at Houston.

* Corresponding author.
E-mail addresses: mefoster@usf.edu (M.E. Foster), JasonAnthony@usf.edu (J.L. Anthony), Douglas.Clements@du.edu (D.H. Clements), Julie.Saram@du.edu (J. Sarama), Jeffrey.Williams@uth.tmc.edu (J.J. Williams).

https://doi.org/10.1016/j.ecresq.2018.01.009

language (L2) phonological awareness and grammatical ability were associated with concurrently measured L2 logical operations skills and numeracy skills in a sample of bilingual Turkish and Moroccan (students' first language – L1) speaking kindergarten students learning Dutch (students' L2). In a study of Canadian students, Romano et al. (2010) demonstrated that English vocabulary in kindergarten predicted third grade English mathematics achievement for students whose L1 was French. These studies suggest that L2 language competence is important to L2 mathematics outcomes in bilingual and DLL students. One would therefore expect English language proficiency at school entry to be predictive of English mathematics outcomes at the end of kindergarten for Hispanic DLLs in the U.S. However, the extent that this relation is a matter of linguistic formatting of test items is not completely clear (see Abedi & Lord, 2001; Rhodes, Branum-Martin, Morris, Romski, & Sevcik, 2015).

2.1. Kindergarten numeracy

Kindergarten number competencies (i.e., understanding of number concepts and number relations; referred to as numeracy throughout this manuscript) are important to children's growth in mathematics achievement (Jordan, Kaplan, Ramineni, & Locuniak, 2009) and are predictive of mathematics learning disability (Mazzacco & Thompson, 2005). Early numeracy also provides a foundation for later academic achievement (Duncan et al., 2007), predicting children's reading achievement better than early literacy skills (Duncan & Magnuson, 2011; Duncan et al., 2007; Koponen, Salmi, Eklund, & Aro, 2013) and predicting mathematics achievement through 15 years of age (Watts, Duncan, Siegler, & Davis-Kean, 2014). Finally, mastery of numeracy concepts allows for stronger understanding of more complex mathematical problems such as applied problem solving (Foster, Anthony, Clements, & Sarama, 2015; Foster, Anthony, Clements, Sarama, & Williams, 2016) and problem solving within the areas of measurement, data analysis, and geometry (National Research Council, 2009).

2.2. Kindergarten mathematics programs

High-quality mathematics instruction helps increase mathematics achievement and helps prevent or mitigate mathematics learning difficulties (Clements & Sarama, 2011; Cross, Woods, & Schweingruber, 2009; Foster et al., 2016; Magnuson, Meyers, Ruhm, & Waldfogel, 2004). Moreover, benefits derived from high-quality mathematics instruction provided during the preschool to early elementary school period appear greatest for children from low-income families and children whose parents have little education (Case, Griffin, & Kelly, 1999; Clements, Sarama, Spitler, Lang, & Wolfe, 2011; Griffin & Case, 1997; National Academy of Sciences, National Academy of Engineering, & Institute of Medicine, 2011; Peisner-Feinberg et al., 2001; Shonkoff & Phillips, 2000). However, most mathematics instruction in kindergarten is inadequate, "teaching" children what they already know (Engel, Claessens, & Finch, 2013). A more useful framework for mathematics instruction would involve targeting fundamental learning goals (e.g., subitizing, counting and arithmetic using subitizing and counting), adjusting instruction according to individual children's progress along learning trajectories, and teaching each fundamental skill to the level of mastery (Clements & Sarama, 2014; National Research Council, 2009; Sarama & Clements, 2009).

Because Hispanic children in the U.S. are less likely than peers to attend preschool (Barnett et al., 2011), kindergarten is a critical period in these children's lives for acquiring mathematical competencies. However, few studies have evaluated the effectiveness of mathematics programs for Hispanic DLLs in the U.S. (Cross et al., 2009). A noteworthy exception was Wang and Woodworth's (2011) evaluation of DreamBox Learning, an online mathematics computer program that supplements face-to-face mathematics instruction. Of the 583 kindergarten and first-grade students in Wang and Woodworth (2011), 87.3% were classified as Hispanic and 80.6% were classified as English learners. The results indicated that children who received supplemental instruction with DreamBox Learning outperformed comparison students who only received face-to-face mathematics instruction on a broad test of mathematics achievement and on a test of measurement and geometry achievement at the end of the school year. These results are promising, suggesting that supplemental computer-based mathematics instruction can improve early mathematical competencies among young Hispanic DLLs.

2.3. Computer-assisted instruction

In the present study, *computer-assisted instruction* (CAI) in mathematics refers to educational software programs that help students learn and apply mathematical concepts and skills (Harskamp, 2014). CAI is most often used as a supplement to teacher-directed classroom instruction (Slavin & Lake, 2008). Advantages of CAI include 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, 2016; Clements & Sarama, 2018; Citation Blinded for Review). However, educators' questions and reasonable concerns about developmental appropriateness, logistics of implementation, compatibility with core curricula, and effectiveness interferes with widespread use of CAI during the early school years (Clements & Sarama, 2003; Cuban, 2001).

Reviews of the scientific literature generally conclude that CAI can provide substantial benefits for children's learning of mathematics (Baroody, Eiland, Purpura, & Reid, 2013; Sarama, 2010, 2018; Cross et al., 2009; Lentz, Seo, & Grunner, 2014; Li & Ma, 2010; Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009; Slavin & Lake, 2008). One review by the National Mathematics Advisory Panel (2008) indicated that CAI applications that are well designed and well implemented can have a positive effect on mathematics performance, and empirical study supports this conclusion (Harskamp, 2014; Moradmand, Datta, & Oakley, 2013; Nusir, Alsmadi, Al-Kabi, & Sharadgah, 2013). A more recent meta-analysis of rigorous studies similarly concludes that there are positive effects of CAI in mathematics when used as a supplement to children's daily classroom instruction (Cheung & Slavin, 2013). Similarly, another meta-analysis of studies examining the use of CAI for early mathematics found a moderate effect size (Harskamp, 2014), whereas still another meta-analysis found positive effects for the use of technological manipulatives with children in preschool and kindergarten (Moyer-Packenham & Westenskow, 2013). Therefore, CAI represents a viable medium to deliver supplemental mathematics instruction.

2.4. Building Blocks mathematics program

The complete Building Blocks (BB) pre-K mathematics program is a comprehensive curriculum that includes a teachers' edition, assessment and resource guides, manipulatives, and a software suite (Clements & Sarama, 2013). The classroom curriculum and software are designed to develop understanding and skill fluency in the domains of numeracy and geometry. A series of empirical studies have supported the effectiveness of the BB program when all its components have been implemented together for use with preschool age children from low-income backgrounds (e.g., Sarama, 2007, 2008; Clements et al., 2011). However, less is known about the effectiveness of the software that accompanies the BB pre-K mathematics program.

The software 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; Sarama, 2007, 2008; Sarama & Clements, 2009). Thus, this study represents not just an evaluation of one particular suite of technology-based activities but a test of the efficacy of software designed on scientific principles.

Positive correlations between usage rates of the BB software and mathematics outcomes by children in the intervention groups in studies by Clements (Clements & Sarama, 2008; Clements et al., 2011) are consistent with hypotheses that the software is an essential active component of the program and that it may mediate impacts of the complete program. However, the research designs of those studies preclude causal statements concerning the efficacy of the BB software. Foster et al. (2016) therefore recently evaluated the effectiveness of the English version of the BB software with monolingual English-speaking kindergartners from low-income backgrounds when it was used in isolation from the comprehensive BB program. Children were randomly assigned to receive either supplemental CAI in numeracy using the BB software or supplemental CAI in phonological awareness using Earobics Step 1 software. Children in the BB software condition evidenced higher posttest scores on tests of numeracy and applied problem solving.

2.5. *Overview of the present study*

2.5.1. *Research question 1: does supplemental use of the Spanish version of the BB software activities for numeracy lead to improved mathematics performance?*

The primary purpose of this study was to determine if the Spanish version of the BB software activities for numeracy was effective when used as a supplement to regular classroom instruction but in isolation from the comprehensive BB program. Research shows that the English version of the BB software leads to improved mathematics achievement when used as a supplement to standard kindergarten mathematics instruction but in isolation from the comprehensive BB program (Foster et al., 2016). One would therefore expect that the Spanish version of these activities, which are based on the same *curriculum research framework* (Clements & Sarama, 2007; Clements, 2007) and model within that framework that details the development of scientifically based software (Clements & Battista, 2000), are effective. Further, the software features not just children's native language (Celedón-Pattichis, 2010; Espada, 2012; Turner, Celedón-Pattichis, & Marshall, 2008), but connects this language to many dynamic visual models (Clements & Sarama, 2014), allows children to create their own problems (Janzen, 2008; Turner et al., 2008), and allows them additional time to do so (Turner & Celedón-Pattichis, 2011). However, the extent to which the Spanish version of the BB software improves mathematics achievement in Hispanic DLLs when it is used as a supplement to standard mathematics instruction has not been investigated. In the present study, kindergarten children, who were Hispanic DLLs, were randomly assigned to one of two experimental conditions: CAI with the Spanish version of BB software or CAI with the Spanish version of Earobics Step 1. All CAI was supplemental to children's daily mathematics and literacy instruction and was provided throughout most of their kindergarten school year. Because we value achievement demonstrated on English and Spanish tests, mathematics achievement was assessed in both languages before and after the intervention. We hypothesized that Hispanic DLLs from low-income backgrounds who received supplemental CAI with the Spanish version of the BB software would demonstrate higher levels of posttest mathematics achievement after controlling for pretest numeracy than children who received supplemental CAI with Earobics Step 1.

2.5.2. *Research question 2: what is the added value of vocabulary proficiency to mathematics outcomes?*

Because language proficiencies arguable influence mathematical development (Figueras-Daniel & Barnett, 2013; Foster et al., 2015; Galindo, 2010; Praet, Titeca, Ceulemans, & Desoete, 2013; Purpura & Ganley, 2014; Purpura et al., 2011), we examined the predictive value of Spanish and English vocabulary on posttest mathematics and if vocabulary proficiencies moderated the effects of the BB software. We hypothesized that vocabulary would predict children's mathematics skills after accounting for autoregressive mathematics ability and intervention effects. We also expected that vocabulary proficiencies would moderate the effectiveness of the BB software on mathematics outcomes.

3. Method

3.1. *Participants*

Participants were recruited from kindergarten classes in five schools in a large urban school district in Texas. Schools were chosen because they served a large proportion of Hispanic DLL students. The percentage of Hispanic students at participating schools ranged from 74% to 99% ($M = 86.4$, $SD = 13.37$). The percentage of students being served as English language learners at each school ranged from 40% to 57% ($M = 49.80$, $SD = 6.38$). The percentage of students eligible for free or reduced lunch ranged from 90% to 98% ($M = 95.8$, $SD = 3.27$).

Across three annual cohorts, 31 kindergarten teachers participated. All teachers were college educated, certified by the state to teach in the public school system, and all but one was female. Racial/ethnic data were reported for 25 of the 31 teachers: 6 were Black, 1 was non-Hispanic White, and 18 were Hispanic. Because three annual cohorts of kindergarten children participated in the study and because some teachers were assigned to the same grade in subsequent years, a few teachers participated in the study more than one year. Specifically, two teachers participated all three years, five teachers participated in two years, and 24 teachers participated in one year, resulting in 40 kindergarten classrooms.

All 40 classrooms of children received full-day kindergarten programming. The average class size was 20 students. On average, 72% of students in participating classrooms were native speakers of Spanish and 28% were native speakers of English. Of the teachers from the 40 classrooms, 32% reported that all of their mathematics instruction was provided in English, while 24% reported that all of their mathematics instruction was provided in Spanish. The remaining 44% of the classrooms reported using English and Spanish to varying degrees. Teachers reported that daily mathematics instruction ranged from 45 to 125 min (M = 77.51, SD = 18.41). Teachers were also asked to report their goals for mathematics instruction and individual teaching opportunities on a 5-point Likert scale that reflected relative proportions of emphasis on *practicing mathematics* versus *understanding mathematics concepts*. Of the classroom teachers, 3% emphasized *practicing mathematics*, 16% emphasized *understanding mathematics concepts*, and 35% emphasized *building conceptual understanding with some practice*. The remaining 46% of the teachers reported equally emphasizing *practicing mathematics and building conceptual understanding*. No classrooms reported emphasizing *practicing mathematics with some focus on mathematics concepts*.

Participating classrooms reflected commonplace instructional practices in kindergarten that balanced discovery learning with explicit instruction. All but one classroom (97.5%) followed a version of the enVisionMath curriculum for kindergarteners. That curriculum includes daily lesson plans organized into focused topics aimed at developing students' conceptual understanding through practice and problem-solving activities. Understandably, teachers were allowed to alter their pace through the curriculum and to supplement their mathematics instruction with materials and activities that were not part of the enVisionMath curriculum. In fact, of higher priority than following the curriculum was adherence to the 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.

All children in participating classrooms were provided some form of supplemental CAI in Spanish. However, only children whose parents provided active informed consent were enrolled in the study. Of the 270 children enrolled in the study, 98% were Hispanic and 2% were mixed ethnicity, including Hispanic. Half of the sample was female, that is, 51%. All families reported speaking Spanish as the primary language in their home. Children ranged in age from 5.00 to 6.30 years (M = 5.70, SD = .29) when they entered the study.

3.2. *Research design and experimental conditions*

3.2.1. *Study design*

Participants were randomized with equal probability from within each classroom to one of two conditions: CAI in mathematics delivered by the BB software (Clements & Sarama, 2013) or CAI in literacy delivered by Earobics Step 1 (Version 1). Thus, the only component of the BB program that was evaluated in this study was the software. Randomization from within classroom at the student level 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. The study design allowed us to identify the impact of the BB software on children's posttest mathematics achievement relative to interacting with Earobics Step 1 software. None of the literacy instruction delivered by Earobics Step 1 explicitly taught numeracy, quantity, geometric or spatial reasoning, or any other obvious mathematical skills. Classroom teachers were blind to children's assignments to CAI conditions because CAI was provided outside of the main classroom during ancillary time for computers. Importantly, using CAI in literacy as the control condition concurrent with CAI in mathematics as the experimental condition guaranteed that the control group did not receive any additional mathematics instruction during the time that the experimental group received CAI in mathematics. The computerized nature of the control condition also ensured that any positive gains associated with the BB software were not due to enhanced computer skills, enhanced attentional abilities, increased motivation, or increased interactions with adults around CAI.

3.2.2. *General CAI procedures*

Children worked individually on computers in their school's computer lab during the ancillary instructional block designated for computer time. Bilingual research assistants were responsible for setting up hardware and software in each school's computer lab. They were also present in the lab to assist with non-instructional aspects of 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.

Children used stereo headphones during CAI to minimize distractions 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 min of CAI per week in addition to the standard instruction they received in their classrooms. The 90 min of CAI per week were delivered in two 45-min sessions; however, because of one school's schedule (eight classrooms), CAI was provided during three 30-min sessions. The duration of 21 weeks of CAI was spread across 30 calendar weeks to accommodate the school district's holidays, district-wide standardized testing, and Kindergarten progress monitoring assessments.

It should be noted that BB software and Earobics Step 1 are both adaptive software programs. Each software program 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 activity becomes more and more difficult until either all levels of the activity are successfully completed or until a child chooses to discontinue that activity by choosing to complete a different activity. Thus, children directed their own instruction insomuch as they were free to move from activity to activity within a given software program and by responding either correctly or incorrectly to each learning trial.

3.2.3. *Computer software programs*

The BB software is appropriate for children in preschool through third grade. It teaches fundamental mathematical ideas through a series of leveled games that comprise a given learning trajectory along two separate strands of 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 given access to interact with all levels of the games that teach numeracy (see online Supplement, Appendix A, for the list of numeracy games and their description).

Earobics Step 1 includes six educational games that teach phonological awareness, short-term memory, sound discrimination, and letter sound correspondence, appropriate for children in preschool through second grade. Participants in this study were given access to interact with the three games that teach phono-

---

[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.

logical awareness (i.e., Caterpillar Connection, Rhyme Time, and Rap-a-Tap-Tap).

3.2.4. Fidelity of implementation

Children's access privileges to a specific software and specific games were programed according to experimental condition. Research assistants ensured CAI was implemented according to study design specifications by monitoring children's usage via daily reports generated online by each software program. Doing so helped ensure that children were moving along their learning trajectories within a software program. As mentioned, 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, district-wide standardized testing, or technical difficulties, such as Internet connectivity. However, individual children or whole classes of children that missed a CAI session made up the missed session within a two-week period.

3.3. Assessment measures and data collection plan

Skills were measured in English and Spanish at four time points. Wave one (i.e., pretest) occurred at the beginning of the school year, preceding CAI by one or two weeks. Numeracy and vocabulary were assessed at wave one. Wave two and wave three test administrations followed 10 and 20 weeks of CAI, respectively. Only vocabulary was assessed at these waves. Wave four (i.e., posttest) occurred toward the end of the school year, following CAI by one or two weeks. Numeracy, vocabulary, and applied problem solving were assessed at wave four.

3.3.1. Vocabulary

Children's vocabulary was estimated using the Expressive One Word Picture Vocabulary Test (EOWPVT; Brownell, 2000a) and the EOWPVT Spanish-Bilingual Edition (EOWPVT–SBED; Brownell, 2000b). Both versions present 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?" Similarly, children were asked to verbally respond to prompts for the EOWPVT-SBED such as "¿Que és esto?", "¿Qué está haciendo ella?" and "¿Qué son estos dibujos?" Standardized administration and scoring procedures were followed for the EOWPVT. However, to permit separate estimates of English vocabulary and Spanish vocabulary, nonstandard Spanish only scoring was employed for the EOWPVT–SBED.[2] Cronbach's alphas demonstrated that both vocabulary measures were internally consistent, .94 and correlations across kindergarten within a given language were high (see online Supplement, Appendix B).

3.3.2. Numeracy

Numeracy skills were assessed in English with the Research-based Early Maths Assessment (REMA; Clements, Sarama, & Liu, 2008), and in Spanish with a forward translation of the REMA. The REMA was chosen because it is validated and designed for early mathematics, its sensitivity to detect differences in the early mathematics performance of young children (Clements et al., 2008) and sensitivity to intervention effects (e.g., Clements and Sarama, 2007, 2008; Clements et al., 2011; Foster et al., 2016). Items from the numeracy strand were administered because they indexed an outcome proximal to the BB software. The numeracy strand includes four subscales: number recognition and subitizing, composition of number, arithmetic, and number comparison and sequencing. Core mathematics skills assessed within the numeracy 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. Standardized administration and scoring procedures were followed. Although the REMA is aligned along the same research-based developmental progressions as the BB software, it assesses numeracy skills more broadly and uses different tasks and materials so that it does not serve as an assessment specific to the BB program. For a detailed description of the REMA's development, see Clements et al. (2008). The internal consistency of the English and Spanish REMA when used with the present sample was good (.84–.90) and stability was evidenced by moderate correlations between pretest and posttest scores within a given language (see Appendix B).

3.3.3. Applied problems

Because of our interest in examining the BB software's impact on distal mathematics outcomes, children were also assessed in English using the Applied Problems subtest of the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2007) and in Spanish using the Problemas Aplicados subtest of the Batería III Woodcock-Muñoz Pruebas de Aprovehamient (Batería III; Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005). These subtests require children to analyze and solve verbally presented problems that involve the application of arithmetic skills to solve mathematics problems. Standardized administration and scoring procedures were followed. Internal consistencies for the English and Spanish administration were good (.86 and .90) when used with the present sample at posttest.

3.3.4. Examiners

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

4. Results

As expected, comparable wave one standard scores[2] of participants' vocabulary suggested that they were stronger in Spanish ($M = 81$, $SD = 17$) than English ($M = 60$, $SD = 12$). Standard scores of participants' wave one nonverbal (or fluid) IQ[3] as assessed with

---

[2] To permit separate estimates of vocabulary and mathematics in English and Spanish, English only scoring was employed for measures of English abilities and Spanish only scoring was employed for measures of Spanish ability. Thus, if a participant responded in English to a Spanish test item or in Spanish to an English item, the student was prompted to respond in the appropriate language. Following the prompt, if the student continued to respond in the incorrect language, the item was scored as incorrect. Of the measures, this procedure deviates from the standardized administration procedures for the Expressive One Word Picture Vocabulary Test – Spanish Bilingual Edition (EOWPVT-SBED). Therefore, the reported standard scores for Spanish vocabulary are likely an underestimate of the sample's norm-referenced achievement.

[3] Because nonverbal (or fluid) IQ is not expected to be dependent on the language of test administration, we only report scores for a forward translation (i.e., from

the Copying and Pattern Analysis subtests from the Stanford-Binet fourth edition (Thorndike, Hagen, & Sattler, 1986) were below average ($M = 79$, $SD = 11$).

4.1. Preanalysis data inspections

4.1.1. Attrition

Of the 270 participants, 16 from the BB software group and 18 from the Earobics Step 1 group relocated after pretesting to a nonparticipating classroom or school (i.e., 13%). These participants and those that remained evidenced equivalent pretest distributions for vocabulary measured in English ($F = .28$, $p = .60$) and Spanish ($F = 1.34$, $p = .25$), nonverbal IQ measured in English ($F = 3.17$, $p = .09$) and Spanish ($F = 1.18$, $p = .28$), and numeracy measured in English ($F = .98$, $p = .32$) and Spanish ($F = .24$, $p = .62$). Therefore, all pretest data were retained for subsequent analyses. Mathematics performance of children is summarized by wave for the full sample and by group in Table 1. In general, children evidenced improvement in numeracy from pretest to posttest and there was no evidence of data non-normality.

4.1.2. Pretest differences

To verify the success of random assignment, we evaluated whether or not the experimental groups differed on demographic characteristics and competencies at pretest. The groups did not demonstrate differences in age ($F = .55$, $p = .46$) or gender ($\chi^2 = .03$, $p = .87$). The experimental groups also evidenced equivalent IQ ($F = 3.53$, $p = .06$), vocabularies measured in English ($F = .35$, $p = .56$) and Spanish ($F = .08$, $p = .78$), and numeracy measured in English ($F = .05$, $p = .82$) and Spanish ($F = .41$, $p = .52$).

4.1.3. Intraclass correlations

Models conditional on pretest numeracy achievement were estimated for the full sample to compute intraclass correlations (ICCs). Conditional ICCs for numeracy and applied problem solving were stronger in magnitude for Spanish (.084, .090) than English (.000, .010). Virtually no variance was attributable to the classroom for English mathematics outcomes when pretest numeracy was included in the models. In contrast, stronger ICCs for Spanish mathematics outcomes suggests that there is less homogeneity in these scores across classrooms. Given the ICCs and desire to maintain consistency across analyses, corrections were applied to standard errors to adjust for classroom level differences when analyzing all outcomes to guard against Type 1 error and biased parameter estimates (cf. Peugh, 2010; Singer & Willett, 2003).

4.2. Data analytic approach

Consistent with Graham's (2009) definitions of missingness, pretest data were missing at random (MAR) by design. Children who scored ≥ 10 items correct on the EOWPVT were administered mathematics assessments in English. Children who scored ≥ 9 items correct on the EOWPVT-SBED were administered mathematics assessments in Spanish. These criterions were chosen based on their similar age equivalents and the amount of language competence presumed to be needed to understand the tasks and directions of the mathematics assessments. Children scoring above both criteria were assessed in both languages at pretest. As a result, 130 children were missing pretest scores for English numeracy ($ns = 62$ and 68 for the BB and the Earobics Step 1groups), while 35 children were missing pretest scores for Spanish numeracy ($ns = 16$ and 19 for the BB and the Earobics Step 1 groups). Thus, there were 140 children with the English REMA at pretest, whereas 235 children completed the Spanish REMA at pretest. For additional details regarding missingness, please see Table 1. To avoid biased estimates and loss of statistical power from listwise deletion of cases missing mathematics data at pretest, we used full information maximum likelihood (FIML) and an auxiliary variable, as recommended by Graham (2009) when data are missing by design. The auxiliary variable is used to explain the missingness and is not included in the model. Therefore in the present study, vocabulary scores at pretest served as auxiliary variables in Mplus (version 7.4) and FIML was used to estimate the regression coefficients using all available data, thereby maintaining the variance structure and not losing cases with incomplete data (Fitzpatrick, McKinnon, Blair, & Willoughby, 2014). FIML estimation is considered as one of the best methods for handling missing data (Muthén, Kaplan, & Hollis, 1987; Worthke, 2000).

Classroom was specified as the cluster-level variable in all analyses to correct for classroom level differences (i.e., level 2) while evaluating the impact of experimental condition on children's mathematics achievement (i.e., level 1). Raw scores were used in all analyses and continuous predictors (i.e., vocabulary) were centered at the grand mean to improve interpretability. Completely standardized results are reported, including R^2 for each model, which is the proportion of variance shared with the optimally weighted independent variables. Effect sizes[4] were estimated according to procedures described by Feingold (2009). Effect sizes reported are for the highest order effect within a given regression model (i.e., the main effect of group or the interaction of pretest numeracy scores and group).

We employed a univariate approach to evaluate impact of the BB software on kindergartners' mathematics achievement because this approach was consistent with our interest in examining impact on proximal (i.e., numeracy) and distal (i.e., applied) mathematics achievement. Impacts were also evaluated separately by language of test administration. There was no evidence of a group by pretest interaction for any of the posttest mathematics scores, so these results are not reported in tables. English numeracy scores from the pretest served as the covariate in the prediction of posttest English mathematics, and Spanish numeracy scores from the pretest served as the covariate in the prediction of posttest Spanish mathematics. To examine the effect of vocabulary on mathematics outcomes, we examined the influence of English vocabulary and its interaction with group in the prediction of English and Spanish posttest mathematics. Similarly, we examined the influence of Spanish vocabulary and its interaction with group in the prediction of Spanish and English posttest mathematics. As mentioned, mean vocabulary, which was derived from wave two and three vocabulary scores, was used for these analyses.

4.3. The prediction of posttest English numeracy

The prediction of posttest English numeracy from pretest English numeracy and group (Model 1) are reported in Table 2. The regression estimates from these models represent the unique effects of each variable adjusting for all other terms in the model. As mentioned above, standard errors corresponding to regression estimates were corrected for classroom nesting. Model 1 indicated that pretest numeracy, but not experimental condition, significantly

---

English to Spanish) of the Copying and Pattern Analysis subtests of the Stanford-Binet, fourth edition (Thorndike et al., 1986).

[4] The following formula adapted from Feingold (2009) was used to determine estimates of effect size: $ES = \beta_{11}\ (time)/SD_{\mathrm{RAW}}$, 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.

Table 1
Descriptive statistics by time point for the full sample and by group.

Variable	Max	Pretest				Posttest			
		n	*Mean*	*SD*	*Skew*	*n*	*Mean*	*SD*	*Skew*
Full sample									
English numeracy	43	140	13.61	5.40	–.03	236	20.08	7.85	–.13
English applied problems	63	–[a]	–[a]	–[a]	–[a]	248	16.34	5.02	–.69
Spanish numeracy	43	235	12.33	6.09	.43	239	20.36	7.85	–.04
Spanish applied problems	63	–[a]	–[a]	–[a]	–[a]	251	17.28	4.63	–.33
Building Blocks group									
English numeracy	43	70	13.71	5.49	.09	116	20.80	8.01	–.19
English applied problems	63	–[a]	–[a]	–[a]	–[a]	121	16.22	5.37	–.81
Spanish numeracy	43	116	12.08	6.35	.57	118	20.88	7.91	–.21
Spanish applied problems	63	–[a]	–[a]	–[a]	–[a]	123	17.70	4.74	–.54
Earobics Step 1 group									
English numeracy	43	70	13.50	5.35	–.17	120	19.39	7.66	–.09
English applied problems	63	–[a]	–[a]	–[a]	–[a]	127	16.46	4.67	–.48
Spanish numeracy	43	118	12.57	5.88	.29	121	19.85	7.79	.12
Spanish applied problems	63	–[a]	–[a]	–[a]	–[a]	128	16.86	4.50	–.14

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

[a] Applied problems was not assessed at pretest.

predicted posttest numeracy. Together, these predictors accounted for 45% ($R^2 = .45$) of the variance in the English numeracy, with an effect size for group of .15.

Results from the prediction of English numeracy at posttest from English vocabulary are reported in Models 2 and 3 (see Table 2). The prediction of English numeracy from English vocabulary was not statistically significant. Expanding Model 2 by including the interaction of group with vocabulary (Model 3) did not result in improved prediction of this outcome. Neither English vocabulary nor its interaction with group, reliably predicted English posttest numeracy.

In contrast to English vocabulary, Spanish vocabulary was a statistically significant predictor of English numeracy at posttest (see Table 2, Model 4). Together, the three predictors accounted for a small although reliable increase in variance of posttest numeracy relative to Model 1 ($\Delta R^2 = .01$). Expanding Model 4 by including the interaction of group with Spanish vocabulary (Model 5) did not result in improved prediction of posttest numeracy. Thus, Spanish vocabulary benefitted children in both experimental groups equally when solving English numeracy problems.

4.4. *The prediction of posttest English applied problems*

Results from the prediction of applied problems from English numeracy and group are reported in Table 2. Although the predictors in Model 1 explained 50% of the variance in applied problems, group was not reliably associated with this outcome. The effect size for group was .14. Thus, the use of the BB software did not correspond to statistically significant or substantively important improvements for English mathematics scores.

After controlling for classroom effects, English numeracy, group, and the prediction of applied problems from English vocabulary were statistically significant (see Table 2, Model 2). These predictors accounted for a small although reliable increase in variance

Table 2
Prediction of posttest English mathematics achievement.

Variable	English numeracy				English applied problems			
	β	*SE*	*p*	R^2	β	*SE*	*p*	R^2
Model 1				.45				.50
Autoregressor	.66	.05	<.001		.71	.05	<.001	
Group	.06	.05	.23		–.05	.04	.22	
Model 2				.45				.52
Autoregressor	.63	.06	<.001		.61	.07	<.001	
Group	.05	.04	.21		–.05	.04	.23	
English vocabulary	.07	.06	.28		.18	.07	.01	
Model 3				.45				.49
Autoregressor	.64	.06	<.001		.61	.07	<.001	
Group	.01	.08	.93		–.08	.08	.36	
English vocabulary	.05	.08	.54		.17	.11	.10	
Group × English vocabulary	.02	.08	.80		.01	.10	.97	
Model 4				.46				.48
Autoregressor	.61	.06	<.001		.70	.05	<.001	
Group	.05	.05	.26		–.06	.04	.18	
Spanish vocabulary	.17	.06	<.01		–.03	.07	.64	
Model 5				.45				.49
Autoregressor	.61	.06	<.001		.71	.05	<.001	
Group	–.02	.12	.90		–.28	.13	.02	
Spanish vocabulary	.15	.08	.06		–.11	.08	.20	
Group × Spanish vocabulary	.07	.13	.60		.26	.15	.08	

Note. Completely standardized results reported.

[a] Autoregressor was English numeracy at pretest.

Table 3
Prediction of posttest Spanish mathematics achievement.

Variable	Spanish numeracy				Spanish applied problems			
	β	*SE*	*p*	R^2	*β*	*SE*	*p*	R^2
Model 1				.42				.44
Autoregressor	.64	.04	<.001		.66	.04	<.001	
Group	.09	.04	.04		.11	.04	<.01	
Model 2				.41				.44
Autoregressor	.64	.04	<.001		.68	.04	<.001	
Group	.09	.04	.05		.11	.04	.02	
English vocabulary	−.03	.06	.62		−.13[b]	.05[b]	.01[b]	
Model 3				.41				.44
Autoregressor	.62	.07	<.001		.68	.05	<.001	
Group	.04	.10	.66		.13	.07	.07	
English vocabulary	−.03	.06	.66		−.10	.07	.16	
Group × English vocabulary	.05	.12	.67		−.04	.09	.63	
Model 4				.50				.55
Autoregressor	.49	.06	<.001		.46	.05	<.001	
Group	.08	.04	.03		.10	.04	.01	
Spanish vocabulary	.32	.07	<.001		.39	.05	<.001	
Model 5				.50				.55
Autoregressor	.50	.06	<.001		.46	.05	<.001	
Group	.26	.13	.05		.12	.15	.43	
Spanish vocabulary	.37	.08	<.001		.39	.05	<.001	
Group × Spanish vocabulary	−.17	.14	.20		−.01	.15	.94	

Note. Completely standardized results reported.
[a] Autoregressor was Spanish numeracy at pretest.
[b] Effects due to suppression and are not trustworthy; see text for details.

in applied problems over that of Model 1 ($\Delta R^2 = .02$). This indicates that English vocabulary uniquely contributed to children's proficiency in solving applied mathematics problems in English. The addition of the group by vocabulary interaction term in the subsequent model (Model 3) did not result in improved prediction of applied problems relative to Model 2. Thus, English vocabulary benefited children in both conditions equally when asked to solve applied problems in English.

Results concerned with the prediction of applied problems in English from Spanish vocabulary suggest that Spanish vocabulary was not significantly associated with this outcome (see Table 2, Model 4). The addition of the group by Spanish vocabulary interaction term (Model 5) did not result in improved predictive ability relative to Model 4. In short, neither Spanish vocabulary nor its interaction with group was reliably associated with improved scores on the test of applied problem solving in English.

4.5. *The prediction of posttest Spanish numeracy*

Both Spanish numeracy at pretest and group significantly predicted Spanish numeracy at posttest, accounting for 42% of its variance (see Table 3, Model 1). The effect size for group was .26. Thus, the use of the BB software was associated with statistically significant and substantively important improvements in Spanish numeracy, after accounting for pretest numeracy and classroom effects.

English vocabulary was not reliably associated with Spanish numeracy at posttest, after accounting for model covariates (see Table 3, Model 2). Expanding Model 2 to include the interaction of group with vocabulary (Model 3) did not result in improved prediction of posttest numeracy. Neither English vocabulary nor its interaction with group was significantly predictive of Spanish numeracy at posttest.

In contrast to English vocabulary, Spanish vocabulary was significantly predictive of Spanish numeracy, after accounting for covariates (see Table 3, Model 4). This model accounted for a moderate increase in variance in posttest numeracy relative to Model 1 ($\Delta R^2 = .08$). The addition of the group by vocabulary interaction term in Model 5 did not account for additional variance in posttest numeracy. Thus, Spanish vocabulary equally benefitted all children when solving Spanish numeracy problems, regardless of experimental condition.

4.6. *The prediction of posttest Spanish applied problems*

Model 1 (see Table 3) indicated that Spanish numeracy and group were significant predictors of posttest applied problem solving in Spanish. Together, the predictors accounted for 44% of the variance in the outcome, with an effect size for group of .31. Thus, the BB software was associated with improvements in Spanish applied problems.

A cursory look at results from analyses including vocabulary in the prediction of applied problem solving for Spanish would seem to suggest that all three terms (English vocabulary, group, and Spanish numeracy) were reliably associated with these scores (see Table 3, Model 2). However, the beta for English vocabulary ($\beta = -.13$) was substantially larger and in the opposite direction than the corresponding zero-order correlation ($r = .04$), indicating that it was inflated by multicollinearity created by introducing the term for the effect of English vocabulary in the model. Consequently, these suppression effects should not be interpreted. Thus, English vocabulary was *not* predictive of higher scores for applied problem solving in Spanish. The addition of the group by English vocabulary interaction term in Model 3 did not account for additional variance in numeracy scores at posttest. Thus, neither English vocabulary nor its interaction with group was significant predictors of Spanish applied problem solving. Spanish vocabulary, however, significantly predicted applied problem solving in Spanish, after accounting for Spanish numeracy and group (Model 4). This model accounted for a moderate increase in variation in applied problem solving in Spanish over that of Model 1 ($\Delta R^2 = .11$). The addition of the group by vocabulary interaction term in Model 5 did not account for additional variation in these scores over that of Model

Table 4
Prediction of posttest Spanish vocabulary.

Variable	Spanish vocabulary			
	β	*SE*	*p*	R^2
Model 1				.85
Autoregressor	.92	.01	<.001	
Group	.01	.02	.69	
Model 2				.85
Autoregressor	.92	.01	<.001	
Group	.03	.09	.76	
Group × Spanish vocabulary	−.02	.10	.80	

Note. Completely standardized results reported.
[a] Autoregressor was Spanish vocabulary at pretest.

4. Thus, Spanish vocabulary benefitted all children equally when solving applied problems in Spanish.

4.7. The prediction of posttest Spanish vocabulary

To examine treatment specificity posthoc and rule out an alternative explanation that participation in the Spanish version of the BB software improved students' tendency to respond in Spanish, as opposed to improving their mathematical knowledge, we ran two additional analyses. However, the use of an auxiliary variable was unnecessary for this analysis because we had complete data at pretest on our measures of vocabulary. As displayed in Table 4, after accounting for the effects of Spanish vocabulary at pretest, group was not reliably associated with Spanish vocabulary at posttest. The addition of the group by vocabulary interaction did not account for additional variation in posttest Spanish vocabulary. Thus, interaction with the BB software improved students' mathematical knowledge, not their tendency to respond in Spanish.

5. Discussion

Identifying effective mathematics interventions for Hispanic kindergartners in the U.S. from low-income backgrounds, especially those learning English and Spanish (i.e., DLLs), represents an important step in reducing mathematics disparities. Increasing the kindergarten mathematics achievement of Hispanic DLLs could yield more promising academic trajectories, higher Science, Technology, Engineering, and Mathematics (STEM) achievement, and potentially increase the number of Hispanic DLLs who later pursue advanced degrees and careers in STEM fields. To this end, we examined the effect of using the BB software as a supplement to children's general education mathematics instruction in schools that served Hispanic DLLs from low-income backgrounds.

The results demonstrated that supplemental implementation of the Spanish version of the BB software used throughout most of the kindergarten school year led to statistically significant and substantively important improvements in children's Spanish mathematics achievement (effect size = .26 and .31 for numeracy and applied problems). These effect sizes exceeded the WWC effect size threshold of .25 that is considered an effect size of substantive importance (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2014). These positive impacts, however, did not generalize to children's performance on measures of English mathematics achievement (effect size = .15 and .14 for numeracy and applied problems). The significant and substantively important impacts of BB software on Spanish mathematics achievement are consistent with the extant research of CAI in mathematics (e.g., Cheung & Slavin, 2013; Harskamp, 2014; Moyer-Packenham & Westenskow, 2013). In addition, it is worth noting that BB software was designed on scientific principles and 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; Clements & Sarama, 2015). However, the only summative evaluation of the BB software was conducted on the English version (Foster et al., 2016). We hypothesized that this software would be equally effective for Hispanic DLLs as their monolingual English-speaking counterparts because it was based on these same scientific principles and because these principles included features particularly useful for DLLs. These features include instruction in children's native language (Celedón-Pattichis, 2010; Espada, 2012; Turner et al., 2008), connections between this language and dynamic visual models (Clements, 2014), opportunities for children to create their own problems (Janzen, 2008; Turner et al., 2008), and the provision of time to do so (Turner, 2011). Thus, this study is the first to validate the Spanish version of the software suite and provides efficacy of the scientific principles for software development.

Results from the present study and the extant research indicate that mathematics learning of children, including Hispanic DLLs in kindergarten, can be enhanced by supplemental use of research-based mathematics software. Nonetheless, it is important to note that small group (non-CAI) numeracy instruction can produce similar effects on kindergarten numeracy achievement as those reported in the present study. Dyson, Jordan, Beliakoff, and Hassinger-Das (2015) demonstrated that a supplemental small group numeracy intervention had substantial effects on children, including English learners, number sense, arithmetic fluency, and mathematics calculation achievement (effect size = .32–.82). The use of teachers in small group instruction is probably an advantage that leads to better outcomes. However, a disadvantage of small group instruction is that it can be teacher intensive. In contrast, supplemental CAI as implemented in the present study is not teacher intensive, does not interrupt children's general classroom mathematics instruction, and does not interfere with other academic programming because it was delivered during students' ancillary block for computer time. Therefore, the effect sizes reported in the present study are of practical significance (i.e., educationally meaningful).

Results of the present study are consistent with our previous evaluation of the English version of the BB software when used with monolingual English-speaking kindergartners (Foster et al., 2016). In the present study, the BB software had similar impacts on scores for the Spanish REMA (i.e., numeracy) and Spanish Applied Problems subtest. The REMA is a measure proximal to the BB curriculum. It was designed to assess all of the progressions in mathematics development that underlie the BB learning trajectories (Sarama & Clements, 2009). Although the REMA is closely aligned conceptually with the BB Software, it assesses numeracy skills using different tasks and different materials so as to not exclusively serve as a curriculum mastery test. Nonetheless, we included Spanish Applied Problems as a distal measure to assess generalization of learned skills. This subtest is not aligned with the BB software and it assesses a broader skill set as a general outcome measure. Demonstration of reliable group differences for solving Spanish Applied Problems therefore strengthens the conclusion that the BB software improved mathematics achievement because the intervention effects generalized to the application of arithmetic skills to solve mathematics problems. Moreover, the effect size for Spanish Applied Problems, .31, was stronger in magnitude than the effect size for the Spanish REMA, .26. Finally, the alternative explanation for the present results that participation in the Spanish version of the BB software improved students' tendency to respond to mathematics problems in Spanish, as opposed to improving their mathematical knowledge, is not tenable given the lack of general impact on Spanish vocabulary and lack of effect for the control group on English mathematics problems.

The differential effect of the Spanish BB software on English verses Spanish mathematics was not surprising. Because the Spanish version of the software was used in the present study, we expected learning gains to be most evident for Spanish mathematics. However, the generalization of mathematical concepts and skills from Spanish to English is an important area of inquiry for Hispanic DLLs in the U.S. The present results suggest that an individual academic school year may be insufficient for Hispanic DLLs to generalize their learning of mathematics from Spanish to English. An alternative explanation is that the linguistic demands of the English mathematics tests exceeded this sample's English proficiency, interfering with their ability to demonstrate mathematics abilities on these tests. Indeed, this sample's achievement on a standardized measure of English vocabulary was more than two standard deviations below the mean of its norming sample and English vocabulary scores were moderately correlated with English mathematics scores. Others (Abedi & Lord, 2001) have demonstrated that decreasing the linguistic demands of English mathematics tests results in higher scores for English learners. Thus, underdeveloped English linguistic competencies or the linguistic demands of English mathematics tests may have interfered with the successful measurement of English mathematics skills in the present study.

In an attempt to identify conditions under which Hispanic DLLs are likely to succeed, we examined the influences of vocabulary in the development of mathematics. English vocabulary was uniquely predictive of solving applied problems in English. Spanish vocabulary was uniquely predictive of solving English numeracy, Spanish numeracy, and Spanish applied problems. These results are consistent with other research that indicates that vocabulary, which is a proxy for language, is involved in solving many different types of mathematics problems (Foster et al., 2015; Hooper et al., 2010; LeFevre et al., 2010; Praet et al., 2013). Further, performance on distal mathematics tests, rather than on proximal tests of numeracy, appear to depend more heavily on children's vocabulary (Foster et al., 2015; Purpura et al., 2011). This is due in part to the linguistic demands associated with testing formats. Indeed, some norm-referenced mathematics tests are characterized by linguistically complex test items (Rhodes et al., 2015; Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006). Language is also the medium through which children connect quantitative knowledge to words and symbols, enabling conceptual development of early mathematics concepts (Krajewski & Schneider, 2009; LeFevre et al., 2010; Purpura et al., 2011). In short, language appears to influence the development of mathematical knowledge and integration of that knowledge with prior learning of mathematics (Purpura & Ganley, 2014).

5.1. Implications

The present study adds to the summative research on the BB program, including our recent evaluation of the English BB software (Foster et al., 2016). Used as a supplement to daily classroom instruction, the Spanish BB software is effective in increasing the mathematics achievement of Hispanic DLLs in kindergarten. The Spanish and English software therefore appear to be an option for school personnel interested in decreasing risk for school failure by increasing the mathematical competencies of kindergartners from minority and low-income backgrounds. In addition, the software program's value added speaks 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 of educational software and consumers. Consumers, administrators, educators, and parents alike desire educational software that teaches multiple competencies and that can be appropriately used with diverse groups of children who vary in school readiness. Thus, use of research-based mathematics software can provide substantial benefits regarding mathematics learning, at least when used as a supplement to general classroom instruction (Citation Blinded for Review).

Given the debate regarding language of instruction in the classroom, the differential effect of the BB software on mathematics outcomes is noteworthy. Learning mathematics in a non-native language may be particularly difficult, especially for kindergartners because they have limited experience in formal education. Children in the present sample also came from homes where Spanish was the primary language spoken. Given the sample's relatively low scores for English vocabulary and because competency in English is the ultimate goal for Hispanic DLLs in some states (e.g., Texas; §89.1201), future research could benefit from evaluating the effectiveness of the English version of the software with these students. Indeed, relative to reading instruction, kindergarten mathematics may prove to be a more ideal starting point for providing Hispanic DLLs with academic instruction in English. However, a mixed presentation of the BB software content, where Spanish instruction and English instruction is alternated daily, may be particularly effective in enhancing kindergarten mathematics achievement of Hispanic DLLs.

Given the impact of the Spanish BB software on Spanish but not English mathematics achievement, and because classrooms in which Hispanic DLLs are taught to employ Spanish and English to varying degrees when providing mathematics instruction, it could be beneficial to assess their mathematical skills in Spanish and English during kindergarten. Doing so could provide a broader and more complete assessment of their mathematical competencies. Such an approach to assessment for Hispanic DLLs may help their teachers identify mathematics concepts in which they are proficient, mathematics concepts in which they need further instruction, and mathematics concepts that need to be generalized from Spanish to English.

To accommodate and support Hispanic DLLs in the classroom, it is important to attend to the relationship between a child's language status and his/her early mathematics competencies (Cross et al., 2009). In particular, English language proficiency may have been a barrier to demonstrating English mathematics proficiency and students in the present study benefitted from Spanish vocabulary when solving English and Spanish mathematics problems. It could therefore be beneficial for mathematics instruction (e.g., teacher led, peer assisted, CAI) for Hispanic DLLs to be supplemented by an intervention that targets language proficiency in English and Spanish, especially vocabulary for specialized mathematical terms and quantitative language. Such an approach could be particularly effective in enhancing the mathematics achievement of Hispanic DLLs. Similarly, future research could examine the tenability of threshold hypotheses (Cummins, 2000), which suggest that certain levels of proficiency in English and Spanish are necessary to facilitate cross-language generalization of mathematics skills.

Future research could extend these studies by examining conditions under which the BB software is effective. For example, the BB software may be differentially effective for Hispanic DLLs in kindergarten based on the quality of their classroom mathematics instruction, the extent to which they receive classroom mathematics instruction in English or Spanish, the extent to which such instruction targets mathematics terminology, and the extent to which their classroom instruction is aligned with the BB software. The present study was not designed to address these questions.

5.2. Limitations

First, the present results may not generalize beyond the population of Hispanic DLLs in Texas kindergartens. For example, the effectiveness of the BB software may differ for other DLLs populations (e.g., Asian) and for Hispanic DLLs who evidence disabilities and as a result need more intensive intervention (see Powell & Fuchs, 2015). Second, the study does not permit statements concerning the merit of the BB software relative to other supplemental mathematics programs. Improved mathematics in the present study may be the result of additional time spent learning mathematics. Other supplemental mathematics programs could be equally beneficial. Third, competence in other aspects of verbal language or more comprehensive measures of verbal language may evidence different relations to mathematics than those reported in the present study. However, recent findings indicate that grammar, vocabulary, and discourse, three facets of language ability, are best represented as a unidimensional latent construct in kindergarten (Language & Reading Consortium, 2015). Given this unidimensionality, it is unlikely that the use of other language measures would have substantially influenced the present results. Fourth, our data do not permit close examination of the influence of mathematics instructional practices employed in children's classrooms, which could have moderated the effect of the BB software. For example, the extent to which children received instruction in English and Spanish may have influenced their mathematics and/or the effect of the BB software on English and Spanish mathematics achievement. Similarly, posttest mathematics may have varied for children receiving English language services relative to those that were not receiving these services. Although classroom level moderators could have occurred, such moderators do not confound the results from the main effects models because classroom instruction was methodologically controlled. Fifth, the majority of classrooms implemented the 90 min of CAI per week as two 45-min sessions. However, in the first year of the project, one school's block schedule (i.e., eight classrooms) resulted in 90 min of CAI being delivered in three 30-min sessions. Nonetheless, the present results are generalizable across administration patterns. Future studies could consider distributed dosage effects, which the present study was not powered to identify.

6. Conclusion

Despite the above limitations, supplemental use of the Spanish and English BB software to improve the mathematical competencies of kindergartners from low-income backgrounds is supported.

Acknowledgements

The research reported in the present manuscript was supported by a grant from the Institute of Education Sciences (IES), U.S. Department of Education (DOE), through Grant R305A080196 "Efficacy of Earobics Step I in English Language Learners and Low SES Minority Children." Preparation of this manuscript was also supported, in part, by another grant from IES/DOE (R324B110007: "Special Education Training Fellowship: Interventions and Professional Development Models in Language and Literacy"). The opinions expressed are solely those of the authors and do not necessarily reflect the position, policy, or endorsement of IES/DOE.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ecresq.2018.01.009.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219–234. http://dx.doi.org/10.1207/S15324818AME1403_2

Anthony, J. L. (2016). For which children of economic disadvantage and in which instructional contexts does Earobics Step 1 improve literacy? *Journal of Research on Educational Effectiveness*, *9*(1), 54–76. http://dx.doi.org/10.1080/19345747.2015.1055637

Barnett, W. S., Carolan, M. E., Fitzgerald, J., & Squires, J. H. (2011). *The state of preschool 2011: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Baroody, A. J., Eiland, M. D., Purpura, 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. http://dx.doi.org/10.3102/0002831212473349

Brownell, R. (2000a). *Expressive one-word picture vocabulary test manual*. Novato, CA: Academic Therapy Publications.

Brownell, R. (2000b). *Expressive one-word picture vocabulary test: Spanish-bilingual edition manual*. Novato, CA: Academic Therapy Publications.

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

Celedón-Pattichis, S. (2010). Implementing reform curriculum: Voicing the experiences of an ESL/mathematics teacher. *Middle Grades Research Journal*, *5*(4), 185.

Cheung, A., & 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. http://dx.doi.org/10.1016/j.edurev.2013.01.001

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: Lawrence Erlbaum Associates.

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. https://www.learntechlib.org/p/17793

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. http://dx.doi.org/10.2307/30034954

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *Journal for Research in Mathematics Education*, *38*(2), 136–163. http://dx.doi.org/10.3102/0002831207312908

Clements, D. H., & Sarama, J. (2010). Technology. In V. Washington, & J. 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. http://science.sciencemag.org/content/333/6045/968.full

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

Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math: The learning trajectories approach* (2nd ed.). New York, NY: Routledge.

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. (2018). 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. http://dx.doi.org/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. http://dx.doi.org/10.5951/jresematheduc.42.2.0127

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

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. http://dx.doi.org/10.17226/12519

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

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. United Kingdom: Multilingual Matters.

Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergartenand first grade (NCES 2002-125)*. Washington, DC: Department of Education.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. http://dx.doi.org/10.1037/0012-1649.43.6.1428

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. http://dx.doi.org/10.5951/jresematheduc.46.3.0331

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

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, *35*(2), 157–178. http://www.jstor.org/stable/43773426

Espada, J. P. (2012). The native language in teaching kindergarten mathematics. *Journal of International Education Research*, *8*(359).

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. http://dx.doi.org/10.1037/a0014699

Figueras-Daniel, A., & Barnett, S. W. (2013). *Preparing young Hispanic dual language learners for knowledge economy*. New Brunswick, NJ: National Institute for Early Education Research. http://nieer.org/policy-issue/policy-brief-preparing-young-hispanic-dual-language-learners-for-a-knowledge-economy

Fitzpatrick, C., McKinnon, R. D., Blair, C. B., & Willoughby, M. T. (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children? *Learning and Instruction*, *30*, 25–31. http://dx.doi.org/10.1016/j.learninstruc.2013.11.003

Foster, M. E., Anthony, J. A., Clements, D. H., & Sarama, J. (2015). Processes in the development of mathematics in children from title 1 schools. *Journal of Experimental Child Psychology*, *140*, 56–73. http://dx.doi.org/10.1016/j.jecp.2015.07.004

Foster, M. E., Anthony, J. A., Clements, D. H., Sarama, J., & Williams, J. M. (2016). Improving mathematics learning of kindergarten students through computer-assisted instruction. *Journal for Research in Mathematics Education*, *47*(3), 206–232. http://dx.doi.org/10.5951/jresematheduc.47.3.0206

Galindo, C. (2010). English language learners' math and reading achievement trajectories in the elementary grades. In E. Garcia, & E. Frede (Eds.), *Young English language learners – Current research and emerging directions for practice and policy* (pp. 42–58). New York: Teachers College Press. http://nieer.org/research-report/english-language-learners-math-and-reading-achievement-trajectories-in-the-elementary-grades-full-technical-report

Graham, J. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. http://dx.doi.org/10.1146/annurev.psych.58.110405.085530

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. (2014). The effects of computer technology on primary school students' mathematics achievement: A meta-analysis. In S. Chinn (Ed.), *The Routledge international handbook of dyscalculia* (pp. 383–392). New York: Routledge.

Hooper, S. R., Roberts, J., Sideris, J., Burchinal, M., & Zeisel, S. (2010). Longitudinal predictors of reading and math trajectories through middle school for African American versus Caucasian students across two samples. *Developmental Psychology*, *46*(5), 1018–1029. http://dx.doi.org/10.1037/a0018877

Janzen, J. (2008). Teaching English language learners in the content areas. *Review of Educational Research*, *78*(4), 1010–1038. http://dx.doi.org/10.3102/0034654308325580

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. http://dx.doi.org/10.1037/a0014939

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., et al. (2016). *The condition of education 2016 (NCES 2016-144)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics. http://nces.ed.gov/pubsearch

Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction*, *19*(6), 513–526. http://dx.doi.org/10.1016/j.learninstruc.2008.10.002

Language & Reading Consortium. (2015). The dimensionality of language ability in young children. *Child Development*, *86*(6), 1948–1965. http://dx.doi.org/10.1111/cdev.12450

LeFevre, J., Fast, L., Skwarchuk, S., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, *81*(6), 1753–1767. http://dx.doi.org/10.1111/j.1467-8624.2010.01508.x

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. http://southernearlychildhood.org/upload/pdf/EarlyTechnology_D42_1.pdf

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. http://dx.doi.org/10.1007/s10648-010-9125-8

Lopez, H., & Velasco, G. (2011). *Childhood poverty among Hispanics sets record, leads nation*. Washington, DC: Pew Hispanic Center. http://www.pewhispanic.org/2011/09/28/childhood-poverty-among-hispanics-sets-record-leads-nation/

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. http://dx.doi.org/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 *Paper presented at the society for information technology and teacher education international conference 2013* http://www.editlib.org/p/48603

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. http://dx.doi.org/10.4018/jvple.2013070103

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Woodcock-Muíoz: Pruebas de Aprovechamiento*. pp. s1. Itasca, IL: Riverside Publishing.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*(3), 431–462. http://dx.doi.org/10.1007/BF02294365

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, DC: U.S. Department of Education. https://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

National Academy of Sciences, National Academy of Engineering, & Institute of Medicine. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. pp. 12984. Washington, DC: The National Academies Press.

National Research Council. (2009). *Mathematics in early childhood: Learning paths toward excellence and equity*. Washington, DC: National Academy Press. https://www.nap.edu/download/12519

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

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. http://dx.doi.org/10.2304/elea.2013.10.3.305

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S., et al. (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. http://www.jstor.org/stable/3654403

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

Powell, S. R., & Fuchs, L. S. (2015). Intensive intervention in mathematics. *Learning Disabilities Research and Practice*, *30*(4), 182–192. http://dx.doi.org/10.1111/ldrp.12087

Praet, M., Titeca, D., Ceulemans, A., & Desoete, A. (2013). Language in the prediction of arithmetics in kindergarten and grade 1. *Learning and Individual Differences*, *27*, 90–96. http://dx.doi.org/10.1016/j.lindif.2013.07.003

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology*, *110*(4), 647–658. http://dx.doi.org/10.1016/j.jecp.2011.07.004

Purpura, D. J., Napoli, A. R., Wehrspann, E. A., & Gold, Z. S. (2016). Causal connections between mathematical language and mathematical knowledge: A dialogic reading intervention. *Journal of Research on Educational Effectiveness*, *10*(1), 116–137. http://dx.doi.org/10.1080/19345747.2016.1204639

Purpura, D. J., & Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology*, *122*, 104–121. http://dx.doi.org/10.1016/j.jecp.2013.12.009

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. http://dx.doi.org/10.1016/j.cogdev.2009.09.003

Rhodes, K. T., Branum-Martin, L., Morris, R. D., Romski, M., & Sevcik, R. A. (2015). Testing math or testing language? The construct validity of the KM-Revised for children with mild intellectual disability and language difficulties. *American Journal on Intellectual and Developmental Disabilities*, *120*(6), 542–568. http://dx.doi.org/10.1352/1944-7558-120.6.542

Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology*, *46*(5), 995–1007. http://dx.doi.org/10.1037/a0018880

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

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*, *11*(2), 105–126. http://dx.doi.org/10.1207/s15326977ea1102_2

Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press. http://dx.doi.org/10.17226/9824

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press. http://dx.doi.org/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 [http://journals.sagepub.com/doi/abs/10.3102/0034654308317473#Texas Education Code, §§7.102(c)(4), 28.002, and 28.0021(a)(1) (2012)].

Turner, E., Celedón-Pattichis, S., & Marshall, M. E. (2008). Cultural and linguistic resources to promote problem solving and mathematical discourse among Hispanic kindergarten students. In R. Kitchen, & E. Silver (Eds.), *Promoting high participation and success in mathematics by Hispanic students: Examining opportunities and probing promising practices [A research monograph of TODOS: Mathematics for ALL]* (Vol. 1) (pp. 19–42). Washington, DC: National Education Association Press.

Turner, E. E., & Celedón-Pattichis, S. (2011). Mathematical problem solving among Latina/o kindergartners: An analysis of opportunities to learn. *Journal of Latinos and Education*, *10*(2), 146–169. http://dx.doi.org/10.1080/15348431.2011.556524

U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. (2014). *What works clearinghouse*. http://ies.ed.gov/ncee/wwc/findwhatworks.aspx

Wang, H., & Woodworth, K. (2011). *Evaluation of Rocketship Education's use of DreamBox Learning's online mathematics program*. Menlo Park, CA: SRI International. http://www.dreambox.com/wp-content/uploads/downloads/pdf/DreamBox_Results_from_SRI_Rocketship_Evaluation.pdf

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

Worthke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Shnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 219–240). Mahwah, NJ: Erlbaum.