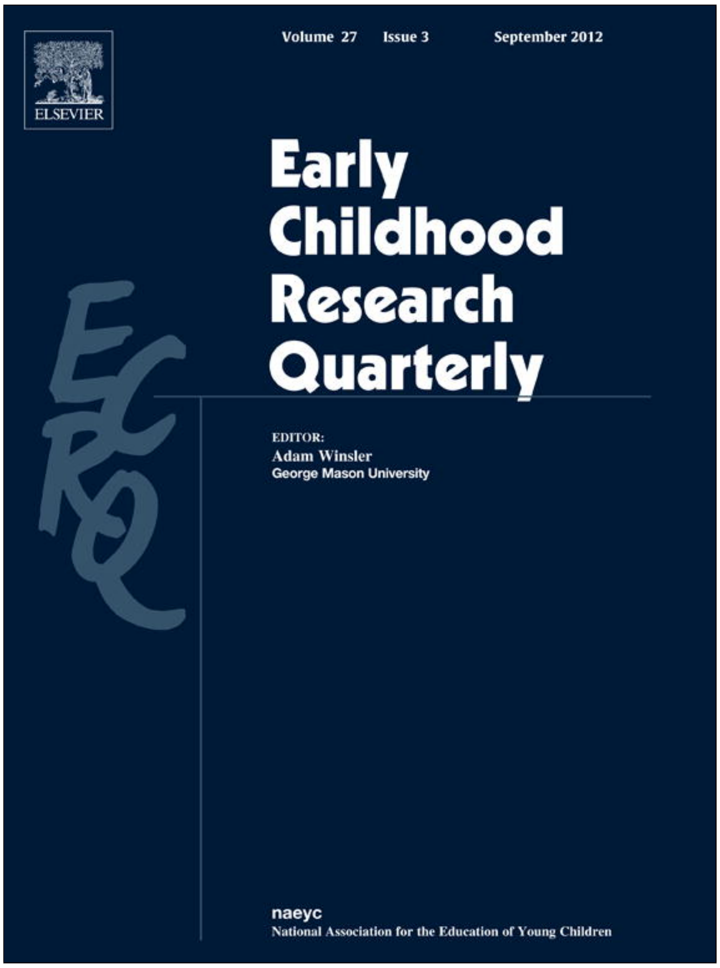
Volume 27 Issue 3 September 2012
ELSEVIER
Early Childhood Research Quarterly
EDITOR:
Adam Winsler
George Mason University
naeyc
National Association for the Education of Young Children

Early Childhood Research Quarterly 27 (2012) 489–502

Contents lists available at SciVerse ScienceDirect

Early Childhood Research Quarterly

The impacts of an early mathematics curriculum on oral language and literacy

Julie Sarama[a], Alissa A. Lange[b], Douglas H. Clements[a,*], Christopher B. Wolfe[c]

[a] University at Buffalo, State University of New York, Department of Learning and Instruction, 505 Baldy Hall, Buffalo, NY 14260, United States
[b] Rutgers University, United States
[c] Indiana University – Kokomo, United States

ARTICLE INFO

Article history:
Received 15 September 2010
Received in revised form
13 November 2011
Accepted 6 December 2011

Keywords:
Preschool
Mathematics
Literacy
Language
Early childhood
Scale-up
Randomized
Curriculum

ABSTRACT

Competence in early mathematics is crucial for later school success. Although research indicates that early mathematics curricula improve children's mathematics skill, such curricula's impacts on oral language and early literacy skills are not known. This project is the first to investigate the effects of an intensive pre-kindergarten mathematics curriculum, *Building Blocks*, on the oral language and letter recognition of children participating in a large-scale cluster randomized trial project. Results showed no evidence that children who were taught mathematics using the curriculum performed differently than control children who received the typical district mathematics instruction on measures of letter recognition, and on two of the oral language (story retell) subtests, sentence length and inferential reasoning (emotive content). However, children in the *Building Blocks* group outperformed children in the control group on four oral language subtests: ability to recall key words, use of complex utterances, willingness to reproduce narratives independently, and inferential reasoning (practical content).

Children who live in poverty and who are members of linguistic and ethnic minority groups demonstrate significantly lower levels of academic achievement, and this pernicious process begins in the earliest years (Denton & West, 2002; National Research Council, 2001; Natriello, McDill, & Pallas, 1990). Preschool education is often provided to address early experiential differences. Unfortunately, many children from lower-resource communities attend preschools that are not of high quality. For example, they are more likely than children from higher-resource communities to be taught by teachers with fewer qualifications (Clifford et al., 2005). This is unfortunate as high-quality programs can help these children achieve greater school readiness upon entry into kindergarten (Clements & Sarama, 2008a; Magnuson, Meyers, Rathbun, & West, 2004; National Research Council, 2001, 2009; Sarama & Clements, 2009a; Shonkoff & Phillips, 2000). In particular, there is evidence that high-quality, research-based preschool mathematics curricula can improve early mathematics achievement (e.g., Clements & Sarama, 2007c, 2008a, 2011a; Sarama, Clements, Starkey, Klein, & Wakeley, 2008). However, with increasing pressure on educators to achieve benchmarks across multiple areas of learning, it is important to know what, if any, impacts these early mathematics programs have on other academic areas, especially language and emergent literacy. This project is the first to investigate the effects of an intensive pre-kindergarten mathematics intervention on the oral language and letter recognition skills of preschool children.

1. Oral language

Both receptive and expressive oral language skills are strongly related to early literacy development (e.g., Cooper, Roth, Speece, & Schatschneider, 2002), later academic success (e.g., Bishop & Edmundson, 1987; Catts, 1993; Pankratz, Plante, Vance, & Insalaco, 2007; Paul, Hernandez, Taylor, & Johnson, 1996; Snow, Barnes, Chandler, Goodman, & Hemphill, 1991), and future linguistic skill (e.g., Conti-Ramsden, Botting, Simkin, & Knox, 2001; Pankratz et al., 2007). Specific components of early oral language, including vocabulary, grammar, semantics, and narrative discourse processes (i.e., complexity and content analysis), have been shown to independently predict later academic success from pre-kindergarten (e.g., Dickinson, McCabe, Anastaspoulos, Peisner-Feinberg, & Poe, 2003). Specifically, early oral language competencies are good predictors of later literacy, nearly equivalent to the established predictors of alphabetic knowledge and phonological awareness (Pearson & Hiebert, 2010).

The NICHD Early Child Care Research Network (2005) demonstrated that these core aspects of oral language skill contributed significantly and independently to later reading success. Further,

* Corresponding author. Permanent address: 114 Carriage Circle, Williamsville, NY 14221-2164, United States. Tel.: +1 716 689 3788; fax: +1 716 645 6721.
E-mail addresses: jsarama@buffalo.edu (J. Sarama), alissa.lange@rutgers.edu (A.A. Lange), clements@buffalo.edu (D.H. Clements), chbwolfe@iuk.edu (C.B. Wolfe).

0885-2006/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ecresq.2011.12.002

oral language was affected by the degree of linguistic complexity in children's environments. Because oral language skills demonstrated a stronger connection to later academic outcomes for children from low-resource backgrounds than children from high-resource backgrounds, rich linguistic experiences at early ages may be especially important for children at risk for academic failure.

2. Letter recognition

Similarly, letter recognition, one early component of orthographic development and pre-reading skill, is a strong independent predictor of later reading success (e.g., decoding, naming speed, phonological awareness, writing, see Denton & West, 2002; McGill-Franzen, 2010; Molfese, Modglin, et al., 2006). For example, letter recognition scores at Kindergarten have been reported as strong predictors of second grade and fourth grade word recognition and reading comprehension (Catts, Fey, Tamblin, & Zhang, 2002).

Letter recognition serves as a base-level step in the process of 'cracking' the alphabetic principle, beginning the process of connecting sound with symbol and growth in phonological awareness (e.g., Adams, 1990; Treiman & Bourassa, 2000; Vellutino, Scanlon, & Tanzman, 1994). Children from low-income backgrounds are at a significant disadvantage in the development of this skill (Bradley & Corwyn, 2002; Denton & West, 2002). For example, in a study of preschool skill gains within a literacy skills-based curriculum with progress monitoring for children from low-income backgrounds, 53% made either no gains or gains of one letter. On the other hand, 47% gained seven or more letters (Molfese, Modglin, et al., 2006). The importance of letter recognition to later literacy development and the wide variability in the growth of letter identification skill suggest the need for highlighting interventions and programs that facilitate the development of this core academic skill (Brown, Molfese, & Molfese, 2008).

3. Linking language and literacy with mathematics

Different bodies of research report conflicting findings concerning the effects of mathematics curricula on early language and literacy. The impact of time-on-task (or time on instruction) on learning provides prima facie justification for the concern of subject-matter conflict (e.g., Bodovski & Farkas, 2007). From this frequently-voiced perspective (see Clements & Sarama, 2009; Farran, Lipsey, Watson, & Hurley, 2007; Lee & Ginsburg, 2007; Sarama & Clements, 2009a), the introduction of a mathematics curriculum could decrease time devoted to language and literacy activities, impeding children's development of those domains. However, this assumes that mathematics activities have little or no positive effects on language and literacy.

Evidence from both educational and psychological research, however, suggests co-mutual beneficial influences. For example, similar developmental milestones exist in the learning of mathematics and language. Children generally begin learning number words at the same time as other linguistic labels. By the age of two, most children recognize which words are reserved for numbers and use these words only in appropriate contexts (Fuson, 1988). Across alphabetic languages, there is a developmental pattern of recognizing word before syllable, syllable before rime-onset, and rime-onset before phoneme (see also Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Ziegler & Goswami, 2005). Similarly in mathematics , development proceeds from conceptualizing numbers as unbreakable quantitative categories to numbers as composites, such as five decomposed into three and two (Butterworth, 2005; Sarama & Clements, 2009a). By the age of six years, children of most cultures have been exposed to both alphabetic and numerical symbol representations and begin to show the ability to segment words into phonemes and numbers into singletons (as in understanding that three is one and one and one, or • • •, Butterworth, 2005; Sarama & Clements, 2009a; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993). Abilities such as identifying the component nature of both words and numbers have been identified as primary predictors of the ability to read (Adams, 1990; Stanovich & Siegel, 1994) and to compute (Geary, 1990, 1993). Finally, deficits in language/literacy and numerosity and competencies are often comorbid among children with learning disabilities (Geary, 1993; Hecht, Torgesen, Wagner, & Raschotte, 2001; Snow, Burns, & Griffin, 1998). Thus, the two domains appear to develop along similar paths.

A second example of possible mutually beneficial influence is the finding that preschoolers' narrative abilities, particularly their ability to convey and relate all the main events of the story and to offer a perspective on the events in the story, predicts mathematics achievement two years later (O'Neill, Pearce, & Pick, 2004). Such abilities may reflect shared relational reasoning competence between narration and mathematics. Further, beginning reading (combined early skills, e.g., letter recognition, word identification, word sounds) is highly predictive of later reading (advanced competencies such as evaluation) only, while beginning mathematics scores are highly predictive of both subsequent reading and mathematics achievement (Duncan et al., 2007). Given the correlational nature of these studies, causal relationships cannot be assumed. However, they suggest that mathematics learning can make a unique contribution to emergent literacy.

Next, consider an ostensibly domain-limited literacy skill such as letter recognition. Recognizing such symbols requires a mental image that distinguishes each symbol from other symbols and graphics. This requires two cognitive competencies, first recognizing (implicitly or explicitly) the components that compose the symbol (e.g., a *b* as a "stick and a circle") and second, recognizing how the components fit together to make an identifiably unique whole (e.g., *b* vs. *p*, Anderson, 2005; Baroody, 1998). These competencies are also integral within early recognition, composition, and decomposition of shapes; competencies that predict later mathematics achievement (Clements, Swaminathan, Hannibal, & Sarama, 1999). Developing these geometric processes within a sufficiently rich mathematics curriculum may therefore serve to support learning across academic domains. Supporting this hypothesis, one geometric program for preschoolers showed positive effects on later measures of literacy and cognition (Razel & Eylon, 1986, 1990). From a similar perspective, Molfese, Beswick, Molnar, and Jacobi-Vessels (2006) reported a significant relationship between low-income preschool children's ability to write numerals and identify letters.

4. The present study

We investigated the effects of a preschool mathematics curriculum on children's learning of language and one measure of emergent literacy. Two cluster randomized trial (CRT) experiments have supported the effectiveness of a research-based early mathematics curriculum, *Building Blocks* in improving mathematics attainment (Clements & Sarama, 2007c, 2008a), and a small-scale "proof of concept" CRT experiment supported the efficacy of the implementation model, TRIAD (Sarama et al., 2008). This study is part of a CRT study of the effects of the curriculum and especially the TRIAD model at a large scale, including distal geographical areas with diverse populations (e.g., Clements, Sarama, Spitler, Lange, & Wolfe, 2011).

We addressed four research questions. The first two test our main hypotheses regarding the effects of a preschool mathematics curriculum, *Building Blocks*, on language and a measure of emergent literacy. The other two questions test our hypotheses about possible moderators and mediators, respectively, for any statistically significant effects.

4.1. What are the impacts of teaching with a preschool mathematics curriculum, Building Blocks, on letter recognition?

The first hypothesis is that the curriculum will strengthen children's ability to recognize letters for two reasons (e.g., Baroody, 1998; Gibson & Levin, 1975). First, the curriculum includes substantial emphasis on geometric/spatial competencies, including spatial skills that may support children's ability to recognize, distinguish, and write letter forms (Clements & Sarama, 2007a). Explicit attention and language within the curriculum focus on identifying basic geometric components, such as line segments and circles, as well as their various combinations and arrangements. For example, geometric decomposition and composition activities include building shapes from components (e.g., line segments, angles, arcs), building mental images of different arrangements of those components (several *Building Blocks* activities were inspired by the aforementioned program, Razel & Eylon, 1986, 1990), and using geometric motions (slides, flips, and turns) to compose geometric shapes to form superordinate shapes (e.g., putting together six equilateral triangles together to form a regular hexagon). In all cases, results are verbally described and compared.

Second, more prosaically, the curriculum gives considerable attention to numerals and therefore visual attributes of alphanumeric symbols (numeral recognition uses processes similar to those used in letter recognition), partially because numerals help children abstract and symbolize mathematical ideas and also because the computer activities require children to read numerals to respond to number tasks (e.g., how many objects are pictured on the screen, Clements & Sarama, 2009).

4.2. What are the impacts of teaching with a preschool mathematics curriculum, Building Blocks, on early oral language?

The second hypothesis is that the mathematics intervention's emphasis on communication, connections between subject-matter domains, representations, problem solving, and reasoning will increase children's oral language competence (Clements & Sarama, 2008a; National Council of Teachers of Mathematics, 2006). Specifically, we expect children learning mathematics through the curriculum to outperform the control group on measures of key word recall, grammatical complexity, independence of narrative retell, and inferential reasoning.

Our rationale for this hypothesis is that one of the curriculum's pedagogical emphases is on children's problem solving and articulation and discussion of their mathematical strategies. Teachers consistently ask, "How do you know?" and "Why?" Children are encouraged to first answer the question by talking with a peer, then share with the small or large group (Clements & Sarama, 2007c). Although preschoolers' initial responses are often general and/or irrelevant (e.g., "I'm smart" or "I thought it in my head"), with modeling from the teacher and peers, most begin to understand and respond with veridical explanations of their cognitive strategies (cf. Ericsson & Simon, 1993). These characteristics, especially the questioning practices, could promote learning beyond mathematics, specifically in language, in four ways. (a) A key characteristic of mathematical thinking and learning in *the curriculum and consensus documents on which it was based* (National Council of Teachers of Mathematics, 2000) is representing and expressing mathematical ideas and situations. To do so, children need to use new concepts and terms (e.g., "angle," "oblique") and use known terms in new ways (e.g., "straight," "share"). (b) Providing mathematics descriptions raises the levels of *precision* of language usage (e.g., "What *is* a triangle?") and often requires a discussion and comparison of different definitions ("looks like an arrowhead" vs. "has three straight sides"), an activity rarely mentioned, even in literacy-rich programs (cf. Preschool Curriculum Evaluation Research Consortium, 2008). Such activity arguably encourages a more complex and thorough processing of the concepts and correlated receptive and expressive language involved in classroom discourse. This also increases sensitivity to the importance of using specific, accurate concepts and vocabulary. (c) Promoting verbal explanations for solutions to problems requires that children be able to explain the cognitive strategies they are using (Lampert & Cobb, 2003). This often requires an increase in grammatical complexity and coherence. (d) Strategy generation itself involves the use and description of reasoning and logical structures, such as categorizations, sequencing, quantification, relationships, comparisons, conditionals, and patterns (Franke, Kazemi, & Battey, 2007; Lampert & Cobb, 2003). These cognitive abilities are foundational to mathematics learning, but they can also be directly related to supporting language—consider how understanding stories such as "The Three Bears" involves all these; for example, simple numeration (three), categories and relationships (size of bears and correspondence between these sizes and sizes of household objects), and ordering and patterns (a sequential plot using a patterned narrative structure).

4.3. Do the variables of gender or ethnic group moderate the relationship between treatment group and language/literacy outcomes?

Research reviews indicate that differences between girls and boys in early mathematics are small and inconsistent (Clements & Sarama, 2009) and curriculum interventions usually report no interactions with gender (Clements & Sarama, 2008a; Preschool Curriculum Evaluation Research Consortium, 2008; Sarama et al., 2008). However, parents' use of spatial language was only related to girls', not boys', mental transformation skill (McGuinness & Morley, 1991) and such spatial language use may be more important for girls (Cannon, Levine, & Huttenlocher, 2007). Thus, it is reasonable to ascertain whether classroom-based mathematics experiences have different effects on girls' and boys' language and emergent literacy. This question is exploratory; however, as gender appears frequently in the literature related to mathematics learning, this analysis will contribute to the field.

Similarly, there has been no consistent evidence of differential effectiveness in most preschool mathematics curriculum interventions for children of different racial/ethnic identities. However, the TRIAD/Building Blocks *intervention was found to be differentially effective for one ethnic/racial comparison: children identified as African-American learned less than other children in the same control classrooms while children identified as African-American learned more than other children in the same* Building Blocks *classrooms* (Sarama & Clements, 2009b). *It may be that the intervention is particularly effective in ameliorating the negative effects of low expectations for learning for children of African-American descent* (cf. National Mathematics Advisory Panel, 2008). If so, it is important to examine the degree to which these resiliency effects could transfer to other academic domains. In a similar vein, effects may be different for other subgroups, such as Hispanic children and, at the school level, schools with different percentages of children with limited English proficiency (LEP) and of those receiving free or reduced school lunch, all of which may moderate any effect of the treatment (National Research Council, 2009).

4.4. If the treatment has significant impacts, are there significant indirect effects through aspects of the classroom and teaching environment on the relationship between children's assignment to treatment group and their mathematics achievement?

The quantity of mathematics activities and the quality of the classroom's mathematics environment, the total number of computers children were using to engage with the intervention's software, and the number of different mathematics activities, significantly mediated mathematics learning (Clements, Sarama, Spitler, et al., 2011) and thus may mediate any relationship between experimental condition and children's development of emergent language and literacy skills. For example, teachers' use of mathematics vocabulary and "math talk" has been related to gains in child mathematics knowledge (Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006). The *Building Blocks* curriculum includes a print curriculum and professional development component in which teachers are explicitly trained in the elicitation of explanations for solutions to mathematical questions. Focus on mathematizing everyday activities in the classroom and the use of explicit techniques to elicit mathematical language and discussion within the classroom may encourage the growth of a classroom climate particularly suited for the development of mathematics and language skills. Aspects of such a climate may include general interactional patterns, but also the number of mathematical activities (including computer activities, which have been shown to generate more academic communication and language than other classroom contexts, see Clements & Sarama, 2003, 2008b).

5. Method

5.1. Participants

The participants used in these analyses were from the first year of the large-scale research project, *Scaling-up TRIAD: Teaching Early Mathematics for Understanding with Trajectories and Technologies*. The two participating school districts were targeted because they traditionally serve children from low-resource communities. To be involved in the study, the districts were required to agree to randomly assign all eligible schools (those whose preschool teachers had not worked in any previous *Building Blocks* project) to treatment groups. District-level adoption was used because the intent was to measure impacts of the curriculum "at scale;" that is, to determine how the curriculum might function in practice when it is adopted district-wide (involving all teachers, not just volunteers). Using a table of random numbers, all eligible schools within each district were publicly (supervised by two school administrators and three staff members) assigned to one of three treatment groups: *Building Blocks*, the *Building Blocks* plus follow through, or the control group. The first two treatment groups were identical for the purposes of the present study. In subsequent years, the *Building Blocks* plus follow-through group was characterized by additional training for kindergarten teachers but this did not occur until after the current study's data were collected and analyzed. Therefore, in the current analysis, only two groups were compared, 26 *Building Blocks* and 17 control schools. Up to 15 children from each class whose parents provided consent were randomly selected to participate.

The present study involved two subsets of the full original sample. The first included all those for whom we have letter recognition scores: 1037 children (50% female; 56% African American, 19% Hispanic, 19% White) from 38 schools (24 *Building Blocks*, 15 control), out of the 1305 children in the original randomly-assigned 42 schools (26 *Building Blocks*, 17 control) across both districts. The second subset consists of 1027 children (52% female; 54% African American, 21% Hispanic, 19% White), who were assessed by the research team on a measure of oral language. Of the total 1,305 children participating in the larger study, approximately 80% were represented on both language and literacy measures.

Table 1 summarizes the demographics of each sample. We had power of .80 to detect effects of $d = .10$ or less with our sample of 42 clusters (schools), and an average of 24 children in each cluster (school).

5.2. Curricula

5.2.1. Treatment math curriculum

Building Blocks (Clements & Sarama, 2007b) is a National Science Foundation-funded mathematics curriculum based on a comprehensive Curriculum Research Framework (Clements, 2007). The curriculum focuses on two main domains of mathematics, number and geometry/spatial skills; woven throughout these domains are subthemes, such as sorting and sequencing, as well as mathematical processes, both general, such as communicating, reasoning, representing, and problem solving and the overarching mathematizing, and specific, such as number or shape composition and patterning. Together, these concepts, skills, and processes were determined to be critical mathematical building blocks.

Building Blocks' instructional approach is finding the mathematics in, and developing mathematics from, children's activity (Clements & Sarama, 2007b). Children are guided to extend and mathematize (i.e., explicate, articulate, and describe) their everyday activities, from block building to art to songs to puzzles, in mathematical language. Thus, the processes of communicating and reasoning, and mathematizing are continually developed through discussions. Teachers ask students to solve problems or tasks, and then ask such questions as "How do you know?," "Why?," and "Can you tell how you figured that out?" Activities include whole group (about 10 min per day), small group (10–15 min once per week for each child, working in groups of four with the teacher), and centers (including a computer center, 5–10 min twice a week for each child). The curriculum includes 30 weeks of instruction; teachers completed from 24 to 30 weeks. More detailed descriptions

Table 1
Demographics of participants with oral language and letter recognition scores.

	Children								Teachers	Schools		
	Total children (males)	Age[a] (*SD*)	Ethnicity						Total teachers	Total schools	SES[b]	ELL[c]
			African American	Hispanic	White	Asian/ Pacific	Native American	Other				
Students with oral language scores												
Building Blocks	726 (346)	64 (4.2)	55%	20%	19%	3%	2%	<1%	71	26	82.4%	11.8%
Control	301 (144)	64 (4.0)	51%	24%	17%	3%	2%	<1%	34	17	82.3%	16.2%
Total	1,027 (490)	63.9 (4.09)	54%	21%	18%	4%	2%	<1%	105	42	82.3%	13.5%
Students with letter recognition scores												
Building Blocks	714 (355)	59.3 (4.07)	58%	17%	21%	3%	1%	<1%	67	24	82.9%	10.9%
Control	323 (161)	59.4 (4.05)	52%	24%	16%	6%	1%	<1%	30	15	84.1%	14.5%
Total	1,037 (516)	59.3 (4.06)	56%	19%	19%	4%	1%	<1%	97	38	83.4%	12.2%

[a] Age in months at time of assessment.
[b] Mean percent free or reduced lunch in schools.
[c] Mean percent English language learning (ELL) in schools.

of *Building Blocks* are available (Clements & Sarama, 2004, 2007b, 2007c, 2011b). CRT evaluations of the curriculum have yielded effect sizes ranging from .50 to 2.10 (Clements & Sarama, 2007c, 2008a).

5.2.2. District's preschool literacy curricula

The first district implemented the Houghton-Mifflin pre-K literacy curriculum, *Where Bright Futures Begin*. The curriculum features ten thematic segments (e.g., *Animals Everywhere*), each consisting of three weeks of theme-related instruction. The flexible lesson structure is designed to develop critical early learning skills, including oral language, listening comprehension, vocabulary, phonological awareness, print awareness, and alphabet knowledge. It also teaches early mathematics skills on a daily basis, with multiple topics taught during small group and whole group times. The teacher manual includes support for teaching English language learners (ELL) and for implementing formative assessment strategies (see www.hmhschool.com). Professional development for *Where Bright Futures Begin* was provided for teachers three times during the year, each time with an emphasis on literacy.

The second district implemented a comprehensive, integrated curriculum, *Opening the World of Learning* (*OWL*), which was designed for full day implementation, with components added to language and literacy, including mathematics, science, social studies, art, and social–emotional development. *OWL* mathematics activities were presented as small-group activities. Components included suggested vocabulary, with procedures provided for extra support as well as extension activities (see www.pearsonlearning.com/microsites/owl/main.cfm). Six professional development sessions on *OWL* were provided during the year.

5.2.3. Fidelity of implementation and effects on mathematics

A previous report (Clements, Sarama, Spitler, et al., 2011) documented that teachers implemented the curriculum with adequate fidelity. On a 5-point Likert scale items, with −2 as strongly disagree and +2 as strongly agree, the mode was 1, and the mean was .77 in fall and .86 in spring. Less than 15% of teachers had an average below .50 (Clements, Sarama, Spitler, et al., 2011, p. 141). Children in the *Building Blocks* group learned more mathematics than the children in the control group (effect size, $g = .72$).

5.3. Measures

5.3.1. Letter recognition

Schools administered assessments of letter recognition at the end of the pre-K year. The assessments were PALS-PreK (Invernizzi, Sullivan, Swank, & Meier, 2004) and MCLASS:CIRCLE (Landry, 2007). Both measures have undergone significant development and piloting, and are widely used as early childhood emergent literacy screeners (Invernizzi et al., 2004; Landry, 2007). PALS-PreK has good inter-rater reliability (.99 across all tasks), split-half reliability (Guttman split-half reliability ranging from .71 to .94 across tasks), and internal consistency (Cronbach's alpha ranging from .75 to .93, Invernizzi et al., 2004). The MCLASS:CIRCLE has good internal consistency (Cronbach's alpha ranging from .90 to .93), high stability of letter recognition over time (ICCs ranging between .71 and .76), and a strong relationship between letter recognition scores and print awareness on the Preschool Comprehensive Test of Phonological and Print Processing (PCTOPPP, Wagner, Torgesen, & Rashotte, 1999) was .76 (Swank et al., 2006). Both assessments measure recognition of upper and lowercase letters by showing letters of the alphabet and asking them to name each letter. The important methodological difference between these two assessments was that the assessments were timed in the first site and untimed in the second. This led to a significant difference in mean scores between the two sites; scores were predictably higher in the second. Therefore, we calculated *z*-scores separately for each site based on the mean and standard deviation of the raw letters correct. These standard scores were utilized in all subsequent analyses.

5.3.2. Oral language

We chose to use The Renfrew Bus Story – North American Edition (RBS; Glasgow & Cowley, 1994), a standardized measure of oral language using narrative retell, to evaluate children's oral language. The assessment involves telling a child a story, and then asking the child to retell the story using the pictures in the wordless storybook as prompts. At the end of the story, assessors asked children two inferential questions.

We assessed oral language in the context of a story retell task because this type of assessment is an ecologically valid method of eliciting language from children as most are familiar with the requirements and procedures of storytelling; most having read stories in school or at home by an early age (Botting, 2002; Curenton, 2004; Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004). The resultant scores from the Bus Story assessment give us indicators of aspects of a child's oral language, such as sentence length or complexity of utterances that are present in the child's story retell.

Previous research has demonstrated strong predictive relationships with literacy and language skills three years after initial assessment (Pankratz et al., 2007). The U.K. version, on which the North American version was based, has extensive validity evidence, including research demonstrating its ability to predict adolescent academic performance from performance on the RBS at four years of age (Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). The North American version is highly correlated with the United Kingdom version on the information (.98) and sentence length (.98) subtests, as reported in the North American version manual (RBS, Glasgow & Cowley, 1994). Test–retest reliability reported in the manual ranged from .58 for complexity to .79 for sentence length (Glasgow & Cowley, 1994). Subscales of the RBS were found (Pankratz et al., 2007) to be strongly related to another measure of oral language, the Oral and Written Language Scales (OWLS, Carrow-Woolfolk, 1996). The RBS sentence length score was significantly related to the OWLS oral composite score ($r = .89$, $p < .0001$), and the OWLS total score ($r = 82$, $p = 0015$), and the information score was significantly related to the OWLS oral composite score ($r = .79$, $p = .0024$, Pankratz et al., 2007).

The stories that children retell during the RBS assessment are transcribed and scored on a series of dimensions by trained coders naïve to the group assignment of the child. The dimensions on which the transcribed stories are scored consist of three primary subtests rating the content of the story retells, including information, complexity, and sentence length. Three additional subtests were two inferential reasoning items (emotive and practical content, respectively) and independence. The validity of the transcribing procedure for the RBS was determined in the present study by evaluating a one-way single measure intraclass correlation (ICC) between a random 7 and 10% sample of stories that were transcribed by two different people (LeBreton & Senter, 2008). The two transcriptions were expert scored by a researcher, and the ICCs between the pairs of scores on the subtests, which represented the difference in transcriptions from the tapes (audio and videotape), were .99 for information, .97 for Sentence length, .94 for complexity.

The information subtest is a measure of how many of the 32 key concepts from the original story the children used in their story retell, with some key concepts being worth two points. The total raw score, with a maximum possible score of 52, is then converted to a standardized score. To score well on this measure, children must remember key concepts (memory), know the meaning of the words representing the concepts well enough to use them appropriately in their retell (vocabulary), and have a sufficient understanding of story structure to use the words or concepts in the right sequence (book/story knowledge). Thus, this subtest represents proficiency for a set of integrated skills. The information subtest is correlated with age (Glasgow & Cowley, 1994), and scores on this dimension have been shown to predict future academic skill (Pankratz et al., 2007).

For this study, we attempted to improve the precision of the measurement by submitting the information items to Rasch analysis. This was the only subtest appropriate for Rasch analysis for two reasons. First, the other subtests include a single item, obviating the use of Rasch analyses. Second, combining subtests would not be desirable as each provides unique information about components of oral language. The Rasch analysis supported a unidimensional construct, although some of the partial credit items, with possible scores of 0, 1, or 2, required recoding. All but three partial-credit items were recoded such that they either received 0 or 1. In these cases, all responses that would have received a 2 were given a 1, and all 0s remained 0s. The final Rasch model for the information scale, with recoded partial credit variables, had an item-reliability of .99 and a person reliability of .79. The slightly lower but acceptable (Bond & Fox, 2001) person reliability may have been due to the overrepresentation of below-average performing children and the limited number of items. The two methods of coding the information scores (i.e., standard score and Rasch scores) were highly correlated ($r = .88$). Inter-rater reliability was calculated using ICC to determine scoring agreement between two different scorers using the same transcriptions (random 10% sample). One-way intraclass correlation with single measures was used to account for error due to having more than two raters, but not always the same two raters rating each target. Agreement was .97 for information.

Complexity is measured by the number of complex utterances children use in their story retells. Complex utterances are defined as those that include a subordinate clause or a relative clause. Complex utterance use is correlated with age (Glasgow & Cowley, 1994). As with information, inter-rater reliability was calculated for a random 10% sample, and was .81 for complexity.

Sentence length is also related to age (Glasgow & Cowley, 1994). This particular measure of sentence length, the mean of the five longest utterances, although correlated with indicators of language development, is a quicker yet less reliable adaptation of the more commonly used measure of oral language utterance length, mean length of utterance (MLU). MLU is not reasonable for use with the RBS because reliable measures of MLU require a minimum of 50–100 utterances (Eisenberg, Fersko, & Lundgren, 2001), and the RBS is intended to be a brief oral language screening tool that often elicits fewer than 50 utterances. Using the same random 10% sample of transcripts as was used for information and complexity, inter-rater agreement was .94 for sentence length.

The inferential reasoning scale includes children's answers to two inferential questions at the end of the assessment. The single, original question from the manual was "Do you think the bus was happy to be on the road again? Why (or Why not)?" This question requires children to make inferences about emotion. A second question that requires inferences that are more practical in nature was added to this administration: "How do you think the driver found the bus?" There are no explicit answers to these questions in the story; therefore, the child must be able to think inferentially to answer the questions. It should be noted that validation on this scale was not reported in the manual, and the scores are based only on two questions. However, as described below, we attempted to bolster what was provided in the manual on this scale by increasing the detail with which the responses were analyzed. This exploratory work may lay the groundwork for future development of the scale.

Children's raw answers to the inferential reasoning questions on the RBS were recoded based on a coding scheme created in addition to the one outlined in the manual. The manual guidelines only provide limited suggestions for how to decide if an answer is "acceptable" or "unacceptable." For example, an answer is deemed acceptable if it is a "moral explanation." We therefore created a second scheme to include a series of scores for each inferential question. We devised the system based on themes that emerged from the data, and subsequently, each response was given a rating comprised of a summary of scores on the following: causal plausibility, reference to story, empathetic, practical, and moral. *Causal plausibility* captured the degree of understanding of causal reasoning in the child's retell. An example of an answer receiving credit on this scale is the following answer to the second inferential question: "The driver found it because he found his tracks." This child found a plausible reason that the driver might have found the bus, when no reason was given in the story. *Reference to story* represented the degree to which the child made reference to the original story in his or her answer to the question. The following answer obtained credit on this scale because the answer refers explicitly to events from the story: "She was running fast, saw bus in water, jumped over gate too." If the child showed evidence of empathizing with the characters in the story, credit was given for the *empathetic* scale. For example, a response that received credit on this scale is one child's response to the first inferential question: "Cause of when he went in the water he was sad." Responses that showed evidence of practical reasoning were given credit on the *practical* scale. A response to the first inferential question given credit on this scale is, "Because he was all clean and didn't want to get all mucky and icky." The child was given credit on the *moral* scale if their responses included references to moral reasoning. For example, the following response to the first inferential question that got credit on the *moral* scale is, "Because now he knows that he can't be naughty again."

Scores on item components one and two range from 0 to 2 points, where "0" means the answer to the inferential question does not possess the given property, "1" means that the answer possess the property to a limited extent, and "2" signifies that the response demonstrates an advanced form of this property. For example, a response coded with a "1" on *reference to story* means that there was an indirect or weak connection to the story evident in the child's response. Components three to five were either scored "0" if the answer did not have that property or "1" if the answer did have that property. The total raw summary scores on this measure ranged between 0 and 5 for question one and 0–6 for question two, with a maximum possible range of 0–7. Two raters rated a random 10% sample of the inferential questions. Rater agreement on this scale was calculated using Cronbach's Kappa because the same two raters rated each item. Agreement was .97 on question one, and .98 on question two.

The independence score is based on the amount of prompting the child needs from the assessor to retell the story. The total prompt score is the summed prompt scores with a maximum of four possible prompts available per picture, and prompts are reverse scored. For example, "no prompt" is given 4 points, whereas the fourth prompt, which is the assessor saying, "The bus. . .," is scored 0 points. Children who retell the story without any need for prompting from the assessor will get the highest independence score, whereas children who need prompts will score lower on independence.

5.3.3. Early mathematics

The "Tools for Early Assessment of Mathematics" (TEAM; Clements, Sarama, & Liu, 2008; Clements, Sarama, & Wolfe, 2011) is a measure of preschool children's mathematical knowledge and skills that features two individual interviews of each child, with explicit protocol, coding, and scoring procedures. A Spanish language version was administered to those children identified by their teachers as English language learners. Less than 5% of the sample received the Spanish language version. All assessment sessions were videotaped. Both the videotapes and record forms were evaluated by trained coders naïve to the group assignment of the child. Assessments were evaluated for item accuracy as well as item solution strategies and error type. Concurrent validity was established with a .86 correlation with a separate research-based instrument, and there was a .89 correlation with the Woodcock Johnson III in pilot testing (Woodcock, McGrew, & Mather, 2001). The assessment was refined in three pilot tests and a Rasch model analysis computed, yielding a reliability of .94 for a similar population of children (Clements et al., 2008).

5.3.4. Classroom observation

"Classroom Observation of Early Mathematics—Environment and Teaching" (COEMET), was created based on a body of research on the characteristics and teaching strategies of effective teachers of early childhood mathematics (Clarke & Clarke, 2004; Clements, Sarama, & DiBiase, 2004; Fraivillig, Murphy, & Fuson, 1999; Galván Carlan, 2000; Galván Carlan & Copley, 2000; Horizon Research Inc, 2001; NAEYC, 1991; Teaching Strategies, 2001). The COEMET measures the quality of the mathematics environment and activities with an observation of three or more hours and is not connected to any specific curriculum. Thus, it allows for intervention-control condition contrasts, no matter what the source of the enacted curriculum. There are 31 items, all but four of which are 5-point Likert scales. An example of one of the three items in the section "Personal Attributes of the Teacher" is, "the teacher appeared to be knowledgeable and confident about mathematics (i.e., demonstrated accurate knowledge of mathematical ideas and procedures, demonstrated knowledge of connections between, or sequences of, mathematical ideas)."

Assessors spent no less than a half-day in the classroom, for example, from before the children arrived until the end of the half-day (e.g., until lunch). All mathematics activities were observed and evaluated, without reference to any printed curriculum. The COEMET has three main sections, classroom elements, classroom culture, and specific mathematics activities (SMA). Assessors completed the first two sections once to reflect their entire observation. They completed a SMA form for each observed mathematics activity, defined as one conducted intentionally by the teacher involving several interactions with one or more children, or set up to develop mathematics knowledge (this would not include, for instance, a single, informal comment). Inter-rater reliability for the COEMET, computed via simultaneous classroom visits by pairs of observers (10% of all observations, with pair memberships rotated) was 88%; 99% of the disagreements were the same polarity (i.e., if one was agree, the other was strongly agree). Coefficient alpha (inter-item correlations) for the two instruments ranged from .95 to .97 in previous research (Clements & Sarama, 2008a; Clements, Sarama, Spitler, et al., 2011). Maximum possible scores for each Likert-based subtest were as follows: classroom culture total score, 45; SMA total score, 95; and verbal interaction scale, 50. The ranges for the remaining subtests were as follows: 8.2–92.5 min, time-on-task; 1.5–14, number of math activities; 0–6, number of computers children were using to engage with the intervention's software.

5.4. Procedure

Schools participating in the TRIAD project were randomly assigned to receive *Building Blocks* training (Building Blocks group), or no training (control group). Pre-kindergarten teachers within these schools were notified of their group assignment in the prior year, and teachers in the *Building Blocks* groups received appropriate training, and taught the *Building Blocks* curriculum, substituting those activities for the district's mathematics activities, while teachers in the control group taught the regular district mathematics curriculum without involvement from the research team. The first year of the project involved teacher training and classroom implementation only. In the second year, children received a pre- and post-assessment on their early mathematics knowledge and skills using the TEAM (Clements, Sarama, & Wolfe, 2011), and teachers continued to receive training and classroom observations.

School-administered letter recognition scores for pre-kindergarten children participating in the TRIAD project were collected in the Spring (end of Year 2). Letter recognition data were not available for four schools, as they did not administer the assessment. Children were assessed on the RBS approximately five months following the TEAM mathematics posttest, in early Fall of their kindergarten year. Oral language data were collected for as many children as possible from the original pre-kindergarten sample, but as these data were collected in the Fall of their kindergarten year, not all children were found within the testing window allotted. We limited the assessment period such that no child could be tested after the end of October; 79% of the children were tested within the first month of school.

Assessors trained especially for the project administered all assessments, except for the letter recognition tests, which were administered by the school. Project assessors were primarily masters-level retired elementary school teachers or graduate students in education with experience working with children. Each assessment involved specialized training including background information on the measure, administration procedures, and practice on administration. Specifically, each assessor was required to complete practice tapes that were coded in-house by senior project staff. Similar training was provided to coders. Assessors and coders needed to achieve a level of item administration or item coding of 98% accuracy or higher to become certified, and those who did not meet the criteria were not selected for these tasks. A trained assessor administered the measures to children individually in an open space (e.g., library, hallway) within the school. All assessments were videotaped to facilitate scoring, and RBS assessments also were audiotaped. Audiotapes served as back up in case of videotape malfunctions, and also aided in the verbatim transcription process when the videotapes were difficult to hear.

6. Results

6.1. Comparability of groups

Table 2 presents descriptive statistics on all measures. The two treatment groups, *Building Blocks* and control, were not significantly different at the beginning of the study on the TEAM pretests, $t\ (1{,}026) = -1.086$, $p = .285$. There were no significant differences on TEAM (math) pretests between those for whom we did and did not collect letter recognition scores, $F\ (1,\ 1{,}303) = .012$, $p = .914$, or between those who did or did not receive the oral language assessment, $F\ (1,\ 1{,}303) = .227$, $p = .634$.

6.2. Hierarchical linear model (HLM)

We used hierarchical linear models (HLM) to analyze the impacts of our treatment on literacy and language outcomes. This method was appropriate because students were clustered within classrooms and classrooms were clustered within schools (Raudenbush & Bryk, 2002). Scores for letter recognition and all oral language subtests were analyzed using the software program

Table 2
Descriptive statistics on child-level outcome scores and teacher-level classroom observation scores by treatment group.

	Type	Building Blocks		Control		Total	
		N	Mean (*SD*)	*N*	Mean (*SD*)	*N*	Mean (*SD*)
Child-level							
Mathematics							
TEAM pretest	Rasch	927	−2.68 (.88)	378	−2.57 (.85)	1.305	−2.65 (.87)
TEAM posttest	Rasch	927	−1.17 (.68)	378	−1.63 (.77)	1.305	−1.30 (.74)
Oral language							
Information	Rasch	726	−.77 (.88)	301	−.99 (.93)	1.027	−.84 (.90)
Complexity	Count	726	1.21 (1.30)	301	1.07 (1.21)	1.027	1.17 (1.28)
Sentence length	SS	726	96.5 (14.31)	301	95.4 (15.50)	1.027	96.1 (14.67)
Independence	Sum	726	39.5 (5.86)	301	37.3 (7.66)	1.027	38.9 (6.51)
Inferential reasoning: Q1	Sum	726	3.02 (1.58)	301	3.01 (1.58)	1.027	3.02 (1.56)
Inferential reasoning: Q2	Sum	726	1.68 (1.25)	301	1.47 (1.21)	1.027	1.62 (1.25)
Print recognition							
Letter Recognition	*z*-score	713	−.02 (1.01)	323	.04 (.98)	1.036	.00 (1.00)
Teacher-level							
COEMET							
Classroom culture total score	Sum	71	36.0 (3.91)	34	31.1 (4.70)	105	34.4 (4.76)
SMA total score	Sum	71	72.4 (4.77)	34	68.2 (6.78)	105	71.0 (5.81)
Verbal interaction scale	Sum	71	38.8 (2.99)	34	35.8 (4.50)	105	37.8 (3.81)
Time-on-task	Count	71	32.4 (15.54)	34	27.2 (12.34)	105	3.7 (14.73)
Total number of math activities	Count	71	7.1 (2.27)	34	4.8 (1.70)	105	6.3 (2.34)
Number of computers for children	Count	71	2.5 (1.19)	34	1.3 (1.30)	105	2.1 (1.33)

Note: 1. Pretest TEAM scores were used as a covariate at the child-level (level-1) and aggregated up to teacher-level (level-2) and school-level (level-3). 2. Oral language measured by the Renfrew Bus Story – North American edition (RBS: Glasgow & Cowley, 1994). 3. Letter recognition measured primarily by PALS (Invernizzi et al., 2004) and MCLASS:CIRCLE (Landry, 2007).

HLM 6 (Raudenbush, Bryk, Cheong, & Congdon, 2006). The three-level hierarchical linear model used school at level-3, classroom at level-2, and child at level-1. The grand-mean centered dichotomous treatment variable (e.g., *Building Blocks* group or not) was entered at level-3, the level of randomization. Individual child-level TEAM mathematics pretest scores were used as a covariate at level-1, and aggregated up to levels 2 and 3 to increase the precision of the estimates (Hedges & Hedberg, 2007). The pretest covariates accounted for a sizable amount of the variance in each of the oral language subscores. For example, the pretest covariates accounted for 10% of the variance in the complexity subscore among children within classrooms at level-1, 22% of the variance among teachers within schools at level-2, and 47% of the variance among schools at level-3.

HLM analyses involved a two-step process for each outcome score. First, a full model was analyzed for each language and literacy outcome including the covariates at each level mentioned above, and all theorized predictors entered at their appropriate levels. Predictors at level-1 (child-level) included dummy codes for Gender (female, 0; male, 1), Disability (without disability, 0; with disability, 1), African American (not African American, 0; African American, 1), Hispanic (not Hispanic, 0; Hispanic, 1), and White (not White, 0; White, 1), level-2 (teacher-level) indicators were limited to the mediators, which consisted of data from the classrooms observations, and these were not entered into the full models, but instead were added to the final models. Finally, level-3 (school-level) predictors were treatment, limited English proficiency percentage, social-economic states (SES, measured by percentage of free or reduced school lunch), and the interaction between treatment and LEP and between treatment and SES. Interactions between treatment and each of the level-1 indicators were also included in each full model.

Second, only those predictors with significant independent contributions from the full models were retained and run in the final model for each outcome. Resultant HLM coefficients signifying the treatment impact in each final model are summarized in Tables 3 and 4. Impacts of other significant predictors are described within each outcome section below, and all results are from HLM analysis, unless otherwise specified.

Effect sizes are given in Hedges *g*, which accounts for treatment groups of different sizes. It is calculated by dividing the individual predictor beta coefficient resulting from the HLM analyses by the pooled standard deviation of each outcome variable. Previous analyses have suggested that effect sizes of .20 and above are relevant to policy when based on measures of academic achievement (Hedges & Hedberg, 2007).

6.2.1. *Mathematics*

The impact of treatment group on mathematics (TEAM) posttest scores was significant, and favored the *Building Blocks* group, with an effect size of .72 (details are available in Clements, Sarama, Spitler, et al., 2011).

6.2.2. *Letter recognition*

There was no significant difference between the Building Blocks group and the control group on the number of letters correctly recognized ($g = -.05$, $p = .743$). Hispanic children scored lower than other children on this outcome ($g = -.19$, $p = .02$), and children with IEPs scored lower than those without ($g = -.26$, $p = .007$). There was no main effect of any other variable or any interaction between any moderators and treatment. As is common in children at this age, few were able to identify all of the presented letters (Molfese, Modglin, et al., 2006).

6.2.3. *Sentence length*

There was no significant difference between the treatment group and control on this outcome ($g = .08$, $p = .23$). There were main effects of school-level SES ($r = -.22$, $p = .03$) and for school-level LEP ($r = -.32$, $p < .001$). The higher the SES (i.e., lower percentage of children in the school receiving free or reduced school lunch) and the lower percentage of LEP in the schools, the higher the scores were on sentence length. There was no interaction with treatment group and this variable.

6.2.4. *Information*

The Building Blocks group significantly outperformed the control group on the information subtest of the RBS, with and an effect size of $g = .29$ ($p < .001$). There was a main effect of gender, with males outperforming females ($g = .15$, $p = .008$), a main effect of whether or not a child was White, with White children scoring higher than others ($g = .26$,

Table 3
Impacts of treatment and moderators on the three primary post-intervention oral language outcome variables from three-level hierarchical linear modeling (HLM) analysis (final model), with treatment group at level-3, and team pretest at all three levels.[a]

Fixed effects	Information			Complexity			Sentence length		
	Coeff	*df*	*t*	Coeff	*df*	*t*	Coeff	*df*	*t*
Unconditional models	−.85	41	−16.60***	1.17	41	24.15**	95.82	41	132.58**
School level									
T1 covariate	.46	38	3.50**	.56	39	4.03**	4.32	37	2.23*
Treatment	.26	38	3.94*	.20	39	2.22*	1.18	37	1.18
SES	−.01	38	−2.13*				−.10	37	−2.23*
LEP							−.16	37	−5.47**
SES – treatment									
Classroom level									
T1 covariate	.57	103	5.04**	.36	103	2.27*	5.48	103	3.29**
Child level									
T1 covariate	.44	1,019	14.19***	.50	1,022	10.31**	3.94	1,020	6.96**
Hispanic									
White	.23	1,019	3.34***						
Gender	.13	1,019	2.74**						
IEP									
	Var	*df*	χ^2	Var	*df*	χ^2	Var	*df*	χ^2
Random effects									
School level	<.01	38	40.63	<.01	39	40.68	.18	37	40.71
Classroom level	.03	62	97.49**	.04	62	81.77*	.60	62	69.16
Child level	97.49			1.40			190.55		

[a] *Note*: 1. Pretest TEAM scores were used as a covariate at the child-level (level-1) and aggregated up to teacher-level (level-2) and school-level (level-3). 2. Interactions between each indicator and treatment were tested also. Only significant indicators that were in at least one final model were included in the table. Coeff: coefficient; *df*: degree of freedom; *t*: *t* value; Var: Variance; χ^2: Chi square; LEP: limited English proficiency; IEP: individualized education plan.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

$p < .001$), and a main effect of SES ($r = -.1$, $p = .039$), such that schools with higher SES had higher information scores. There was no interaction between any of these moderators and treatment.

6.2.5. *Complexity*

Children in the Building Blocks group significantly outperformed those in the control group on number of complex utterances used in the story retell ($g = .16$, $p = .03$). The mean number of complex utterances for the Building Blocks group was 1.21 (SD = 1.30),

Table 4
Impacts of treatment and moderators on additional post-intervention outcome variables from three-level hierarchical linear modeling (HLM) analysis, with treatment group at level-3, and team pretest at all three levels.[a]

Fixed effects	Independence			Inferential Q1			Inferential Q2			Letter recognition		
	Coeff	*df*	*t*	Coeff	*df*	*t*	Coeff	*df*	*t*	Coeff	*df*	*t*
Unconditional models	38.80	41	120.46**	3.02	41	61.27**	1.62	41	35.81**	−.012	37	−.18
School level												
T1 covariate	.34	39	.40	.42	39	2.59*	.37	37	2.24*	.45	35	2.20*
Treatment	2.34	39	4.05**	.06	39	.55	.21	37	2.40*	−.05	35	−.33
SES							<.01	37	.69			
LEP												
SES–Treatment							−.021	37	−3.68**			
Classroom level												
T1 covariate	1.13	103	1.31	.20	103	1.08	.55	103	3.89**	.32	95	1.81†
Child level												
T1 covariate	1.27	1,020	4.97**	.27	1,022	4.16**	.312	1,020	6.349**	.45	1,029	12.97**
Hispanic										−.19	1,029	−2.30*
White												
Gender	.86	1,020	2.16*									
IEP										−.26	1,029	−2.72**
	Var	*df*	χ^2	Var	*df*	χ^2	Var	*df*	χ^2	Var	*df*	χ^2
Random effects												
School level	.67	39	57.44*	<.01	39	36.85	<.01	37	38.99	.06	35	69.28**
Classroom level	1.31	62	83.00	<.01	62	59.79	<.01	62	52.69	.13	58	182.46**
Child level	38.26			2.42			1.43			.67		

[a] *Note*: 1. Pretest TEAM scores were used as a covariate at the child-level (level-1) and aggregated up to teacher-level (level-2) and school-level (level-3). 2. Interactions between each indicator and treatment were tested also. Only significant indicators that were in at least one final model were included in the table. Coeff: coefficient; *df*: degree of freedom; *t*: *t* value; Var: Variance; χ^2: Chi square; LEP: limited English proficiency; IEP: individualized education plan.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 5
Mediational impacts of classroom variables on the significant post-intervention outcomes[a] from three-level hierarchical linear modeling (HLM) analysis (final model), with treatment group at level-3.

	Information			Complexity			Independence		
	Coeff	*df*	*t*	Coeff	*df*	*t*	Coeff	*df*	*t*
School level									
T1 covariate	.42	38	3.17**	.47	39	3.37**	.10	39	.12
Treatment	.21	38	2.50*	.06	39	.52	1.89	39	2.95**
Classroom level									
T1 covariate	.55	100	4.88***	.35	101	2.22*	.98	102	1.14
Classroom culture	.01	100	1.52	.01	101	1.49	.09	102	1.46
Number of math activities	<.01	100	.17						
Number of computers	−.01	100	−.30	.05	101	1.56			
Child level									
T1 covariate	.44	1,016	14.18***	.501	1,020	10.31***	1.27	1,019	4.97***
	Var	*df*	χ^2	Var	*df*	χ^2	Var	*df*	χ^2
Random effects									
School level	<.00	38	38.89	<.01	39	36.89	.61	39	56.62*
Classroom level	.03	59	97.78	.03	60	82.58*	1.23	61	81.96*
Child level	.56			1.40			38.26		

[a] *Note*: 1. The only outcomes included were those that were significantly different by treatment and for which mediational impacts were found. The mediational impacts were determined within the final models of each outcome. 2. Pretest TEAM scores were used as a covariate at the child-level (level-1) and aggregated up to teacher-level (level-2) and school-level (level-3). Coeff: coefficient; *df*: degree of freedom; *t*: *t* value; Var: Variance; χ^2: Chi square.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

and the mean for the control group was 1.07 (SD = 1.21). There were no main effects or interactions between treatment group and any of the other predictors on this outcome.

6.2.6. Independence

Children in the *Building Blocks* curriculum needed significantly fewer prompts to retell their stories than children in the control group ($g = .36$, $p < .001$). The mean independence score was 39.5 (SD = 5.86) for the Building Blocks group and 37.3 (SD = 7.66) for the control group. There was a main effect of gender, with males needing fewer prompts than females ($g = .13$, $p = .04$). No main effects of the other predictors were found, nor were any interactions of these variables with treatment significant.

6.2.7. Inferential reasoning

The first inferential reasoning question was the emotive, "Do you think the bus was happy to be on the road again?" Children in the *Building Blocks* classrooms did not differ from the control group in terms of "acceptability" of their response as measured by a Chi Square ($\chi^2 = .22$, $p = .64$), or on total score using the more detailed scoring scheme described in Section 5 ($g = .03$, $p = .61$). In addition, there were no main effects or interactions involving the other predictor variables.

The second inferential reasoning question, was the practical, "How do you think the driver found the bus?" Children in the *Building Blocks* classrooms did not differ from children in the control group in terms of "acceptability" of their response ($\chi^2 = 2.59$, $p = .11$). However, children in the *Building Blocks* groups had significantly higher scores using the detailed coding scheme ($g = .16$, $p = .03$). The mean score for the *Building Blocks* group was 1.7 (*SD* = 1.3), and for the control group was 1.5 (*SD* = 1.21). There were no main effects for the predictor variables, but there was a significant interaction between school-level SES and treatment such that there was a larger treatment group difference for schools with lower SES than for those with higher SES ($r = .24$).

6.2.8. Mediation

The mediational impacts of the classroom environment was conducted utilizing the COEMET classroom observation subtests: classroom culture total score, SMA total score, verbal interaction scale, time-on-task, total number of math activities, and number of computers children were using to engage with the intervention's software. We followed the Baron and Kenny (1986) approach within a multilevel model (Krull & MacKinnon, 1999). First, we calculated whether the treatment variable predicted each COEMET indicator. Then, we calculated whether each of these predicted each of the language outcomes. Note that we only conducted this analysis for those language outcomes that were significantly impacted by treatment. Those variables that predicted the language outcome were deemed to be significant mediators. The magnitude of the mediating impact was determined by comparing the difference in the treatment coefficient with and without the mediators in the model.

The impact of treatment on the information outcome of the RBS was partially mediated by classroom culture, number of computers children were using to engage with the intervention's software, and the number of math activities, with these indicators together mediating 28% of the treatment effect. The classroom culture and the number of computers children were using to engage with the intervention's software mediated the impact of treatment on complexity (69%), whereas classroom culture alone mediated the impact of treatment on independence (19%). None of the other variables mediated any of the impacts. Also, none of the COEMET scores mediated the inferential reasoning impacts found for question two. Results of the mediational analyses can be found in Table 5. (COEMET data indicate that teachers in the treatment group spent about 5 more minutes per day on mathematics than control teachers, although this difference was not statistically significant Clements, Sarama, Spitler, et al., 2011.)

7. Discussion

Certain early mathematics curricula have been shown to improve preschooler's mathematics competences (e.g., Clements & Sarama, 2007c, 2008a, 2011a; Griffin, 2004). The literature includes both theoretical and empirical works that suggest such curricula may also facilitate children's development of language and literacy competencies. For example, mathematics and language appear to have co-mutual influences. The mathematics curricula may benefit emerging literacy due to their emphasis on reasoning, problem solving, and communication (National Council of Teachers of Mathematics, 2006; Senk & Thompson, 2003), the developmental of metalinguistic awareness required by transitions between registers and conventions (Boero, Douek, & Ferrari, 2008), and an emphasis on geometric forms. In counterposition, some educators have expressed concern that increasing mathematics achievement by introducing mathematics curricula will come at the expense of emerging literacy or language skills (see, e.g., Clements & Sarama, 2009; Farran et al., 2007; Lee & Ginsburg, 2007; Sarama & Clements, 2009a). We investigated the effects of an intensive early mathematics curriculum, *Building Blocks*, on the letter recognition and oral language skills of preschool children participating in a large-scale cluster randomized trial. These results do not support the contention that the use of an intensive mathematics curriculum negatively impacts developing literacy or language competencies. Instead, children taught by teachers using *Building Blocks* outperformed children in the control group on four of the oral language subtests and did not differ statistically on the remaining measures.

One hypothesis was that exposure to the *Building Blocks* curriculum during preschool would serve to strengthen a child's ability to recognize letters. Our results suggest, however, that children did not differ on the number of letters identified across research groups. Although we cannot support the added benefit of additional training in geometry or numerals as a supportive factor in letter recognition (cf. Baroody, 1998), these results do not support the contention that the implementation of an intense, trajectory-based mathematics curriculum detracts from growth in this core predictor of later literacy development. Procedurally-based literacy curricula have also had difficulty supporting growth in letter recognition across samples of children from low-income backgrounds (e.g., Brown et al., 2008). Future research is needed to determine whether instruction in geometry and numerals compensated for less time on literacy-based letter instruction, or the reallocation of time did not affect children's learning.

Children identified as Hispanic scored lower than other groups on letter recognition across both *Building Blocks* and control conditions. This finding may reflect that many of these children were learning English as a second language and may not demonstrate the same linguistic proficiency in English as their native-English speaking peers. However, we do not have the information on individual children's English proficiency necessary to confirm this (district data was unavailable for some children, and the type of data was not consistent; e.g., bilingual, limited English proficiency, language preference at home). However, there were no impacts on the other language measures, and no interaction with treatment group, so Hispanic children appear to be able to learn mathematics, literacy, and language through the curriculum and continue to develop as they would have without this instruction. No other ethnic group variables moderated the relationship between treatment group and letter recognition.

Our second major research hypothesis was that the curriculum's emphasis on mathematical communication would affect children's oral language scores. The *Building Blocks* group did demonstrate greater oral language skill as compared to the control condition on four of the six oral language measures: information, complexity, independence, and inferential questions (the practical question). *Building Blocks* children were able to remember and appropriately use key words from the story, used more complex utterances in their story retell, needed fewer prompts from the assessor to retell the story, and provided more complex responses to an inferential question centered on the practical aspects of the story than children in the control condition.

These advantages are of particular note for two important reasons. First, the measure of language, the *Renfrew Bus Story* (RBS), was administered four months after the end of the intervention with a summer break in between, so the difference is not only significant statistically and practically, but has the additional quality of persistence. Second, the primary purpose of *Building Blocks* is to cultivate mathematical knowledge, not develop oral language. Thus, the language assessment was a far-transfer task. Given the distal nature of the oral language assessment and dominant focus on mathematical concepts, this is strong evidence that children acquire competencies through the curriculum that both persist and extend to other areas of learning.

One explanation for the treatment advantage is that the curriculum has an explicit focus on children providing verbal explanations for solutions to mathematics problems. For example, when a child identifies a square out of a set of shapes, a *Building Blocks* teacher demonstrating high curricular fidelity would then ask, "How do you know it's a square?" At the beginning of the school year, young children often answer such a question with simplistic or unrelated responses, such as, "Because it looks like a square" or "I thought it in my head." As the curriculum progresses and the child experiences repeated invitation to explain his/her thought process, supporting even young children's ability to give accurate, reasoned responses, such as, "Because it has four sides and all square corners." O'Neill et al. (2004) argue that prediction of later mathematical ability by certain aspects of preschoolers' storytelling may be explained by the shared relational reasoning ability. The present study's positive effect on inferential reasoning in a narrative context supports that explanation and provides causal evidence that early mathematics can support the learning of language competencies.

Our third research question addressed the moderation of these effects. Only one indictor significantly moderated the impact of only one of the language or literacy outcomes. The treatment had a greater effect within schools with lower average SES than for those with higher average SES. It may be that the curriculum of lower-SES schools include fewer practical inferential reasoning questions; however, this is a post hoc hypothesis and must be checked in future studies. In summary, with that sole exception, there was no evidence that the treatment was more or less effective for any of the other analyzed subpopulations. That is, results were similar across children's gender, ethnic/racial group, and schools' percentage of LEP students.

Our fourth hypothesis addressed mediators of significant treatment impacts. Our results suggested that the general classroom environment, including aspects of the linguistic interactions between teacher and student, is improved when teachers use the curriculum. The classroom culture score within the COEMET observational instrument partially mediated the relationship between treatment scores and outcomes on the oral language measure related to information, complexity, and independence. This section of the COEMET is designed to measure change within the classroom environment reflective of both adherence to the curriculum and an individual teachers use of learning trajectories within early mathematics instruction. The content and frequency of verbally based mathematic interactions with students is a significant component in this evaluation. Our results suggest that components of the curriculum appear to be leading to changes in the general classroom environment, and this contributes, at least in part, to the improved language performance. These components include daily discussions of mathematical concepts and vocabulary, but perhaps more significant, an emphasis on mathematical processes, including problem solving, reasoning, representing, making connections, and communicating. Also significant may be the combination of physically and verbally active engagement with ideas in whole group ("turn to your partner and explain what the answer is and why") and especially small group contexts (cf. Clements & Sarama, 2008a; Klein, Starkey, Clements, Sarama, & Iyer, 2008), which have been correlated with oral language growth (Smith & Dickinson, 1994).

In classrooms where children are explicitly provided the opportunity to explain and discuss their thinking orally, children become more confident in their thinking and verbal expressions (Lappan & Schram, 1989). Indeed, improvements in the independence measure (i.e., fewer prompts children needed to complete the RBS) suggest that children within the *Building Blocks* group may be developing more confidence to verbally express their thinking than children in control classrooms. This confidence in oral expression of mathematical concepts may facilitate transfer to other academic domains. For example, describing, explaining, justifying, and summarizing are communication processes important in both early mathematics and early science. Competence in these processes in early science is related to the development of scientific knowledge (Norris & Phillips, 2003; Yore & Hand, 2003) as well as to children's self-confidence (Mantzicopoulos, Patrick, & Samarapungavan, 2008)

Finally, we found that the number of computers children were using to engage in the curriculum software mediated the impact of treatment on the number of complex utterances the children used in the stories they retold during the oral language assessment. Previous work has shown that the computer activities used in this curriculum generated more academic communication and language than other classroom contexts (Clements & Sarama, 2003, 2008b). The computer-based activities may promote discussions among students. Future research is needed to analyze closely children's interactions as they work on this software.

7.1. Limitations

These data are limited because only posttests for oral language were available. However, using pretests, even if not from the same assessment or topic area, can improve precision in HLM analyses (Bloom, Richburg-Hayes, & Black, 2007). Indeed, the mathematics pretest scores used in the present study did account for a significant amount of variance in the language and literacy outcomes among children. Still, because we did not have pretests for the language/literacy measures, the results would benefit from replication including a baseline measurement of these skills. Although we cannot completely describe the specific emergent literacy activities happening in the classrooms, both districts implemented district-wide curricula programs and participating schools were blocked on district then randomly assigned to condition. Therefore, emergent literacy curriculum activities can be assumed similar across treatment groups.

All children who completed the assessments were included in our sample, including children for whom English is a second language (ESL) and English language learners. To ensure assessments are fair and culturally sensitive, different assessments may be given to children whose native language is not English. In the case of the large-scale study from which these data are gleaned, we were interested in the impact of our mathematics curriculum on children in school districts generally. As the encompassing research was a scale-up study, the focus was not on individual child characteristics, but rather on impacts of large-scale curricula implementation – district-wide. However, we did enter into our HLM analyses language variables such as school-level LEP status, to measure impacts of varying levels of English proficiency in school environments, and did not find a differential effect of treatment by language proficiency environment.

The data on emergent literacy were limited. For print recognition, we only had scores for letter recognition, and recognize that even alongside the oral language measure, our data do not cover all aspects of emergent literacy. Future work would benefit from more comprehensive measures of the spectrum of emergent literacy. For example, we cannot speak to the impact of *Building Blocks* on phonological awareness. We would also like to assess the influence of other teacher-level and parent-level mediational variables in future work. For example, the impacts of teachers' years of education and parents' behavior such as reading with children should also be investigated as possible influences on the relationship between early mathematics and language/literacy development.

7.2. Conclusion

There was no evidence in this study that teaching with a comprehensive early mathematics curriculum will negatively impact letter recognition or language skills of children from low-resource, urban communities. In contrast, a mathematics curriculum can have a positive effect on several critical oral language competencies.

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grants to the University at Buffalo, State University of New York: Grant No. R305K05157, "Scaling Up TRIAD: Teaching Early Mathematics for Understanding with Trajectories and Technologies" and Grant No. R305A070468A, "Comprehensive Postdoctoral Training in Scientific Education Research." The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education. It should be noted that the curriculum evaluated in this research has been published by two of the authors (Sarama and Clements), who thus could have a vested interest in the results. However, author Lange and several external researchers oversaw the research design, data collection, and analyses. Further, after this research was reported, Lange undertook responsibility to revise and disseminate the Renfrew Bus Story – North American Edition instrument, which is used in this study.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Anderson, J. R. (2005). *Cognitive psychology and its implications* (6th ed.). New York, NY: Worth.

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly*, *38*(4), 470–487.

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

Baroody, A. J. (1998). *Fostering children's mathematical power: An investigative approach to K-8 mathematics instruction*. Mahwah. NJ: Erlbaum.

Bishop, D. M., & Edmundson, A. (1987). Language-impaired four-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, *52*, 156–173.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation & Policy Analysis*, *29*(1), 30–59.

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.

Boero, P., Douek, N., & Ferrari, P. L. (2008). Developing mastery of natural language: Approaches to some theoretical aspects of mathematics. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed., pp. 262–295). New York, NY: Routledge.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching & Therapy*, *18*, 1–22.

Bradley, R., & Corwyn, R. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–399.

Brown, T. E., Molfese, V. J., & Molfese, P. (2008). Preschool student learning in literacy and mathematics: Impact of teacher experience, qualification, and beliefs on an at-risk sample. *Journal of Education for Students Placed at Risk*, *13*, 106–126.

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, *46*, 3–18.

Cannon, J., Levine, S. C., & Huttenlocher, J. (2007, March). *Sex differences in the relation between early puzzle play and mental transformation skill*. Paper presented at the Biennial Meeting of the Society of Research in Child Development, Boston, MA.

Carrow-Woolfolk, E. (1996). *Oral and written language scales: Written expression scale manual*. Circle Pines. MN: American Guidance.
Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research*, *36*, 948–958.
Catts, H. W., Fey, M. E., Tamblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research*, *45*, 1142–1157.
Clarke, D. M., & Clarke, B. A. (2004). Mathematics teaching in grades K-2: Painting a picture of challenging, supportive, and effective classrooms. In R. N. Rubenstein, & G. W. Bright (Eds.), *Perspectives on the teaching of mathematics (66th yearbook)* (pp. 67–81). Reston, VA: National Council of Teachers of Mathematics.
Clements, D. H. (2007). Curriculum research: Toward a framework for 'research-based curricula'. *Journal for Research in Mathematics Education.*, *38*, 35–70.
Clements, D. H., & Sarama, J. (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. (2004). *Building Blocks* for early childhood mathematics. *Early Childhood Research Quarterly*, *19*, 181–189.
Clements, D. H., & Sarama, J. (2007a). *Building Blocks Curriculum, Grade PreK*. Columbus, OH: SRA/McGraw-Hill.
Clements, D. H., & Sarama, J. (2007b). *Building Blocks—SRA Real Math Teacher's Edition, Grade PreK*. Columbus, OH: SRA/McGraw-Hill.
Clements, D. H., & Sarama, J. (2007c). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education*, *38*, 136–163.
Clements, D. H., & Sarama, J. (2008a). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*, 443–494.
Clements, D. H., & Sarama, J. (2008b). Mathematics and technology: Supporting learning for students and teachers. In O. N. Saracho, & B. Spodek (Eds.), *Contemporary perspectives on science and technology in early childhood education* (pp. 127–147). Charlotte, NC: Information Age.
Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York, NY: Routledge.
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. (2011b). *TRIAD (Technology-enhanced, research-based instruction, assessment, and professional development) Scale-up Model*. Retrieved from http://UBBuildingBlocks.org
Clements, D. H., Sarama, J., & DiBiase, A.-M. (2004). *Engaging young children in mathematics: Standards for early childhood mathematics education*. Mahwah, NJ: Erlbaum.
Clements, D. H., Sarama, J., & Liu, X. (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.
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.
Clements, D. H., Sarama, J., & Wolfe, C. B. (2011). *TEAM—Tools for Early Assessment in Mathematics*. Columbus, OH: McGraw-Hill Education.
Clements, D. H., Swaminathan, S., Hannibal, M. A. Z., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education*, *30*, 192–212.
Clifford, R., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., et al. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science*, *9*, 126–143.
Conti-Ramsden, G., Botting, N., Simkin, Z., & Knox, E. (2001). Follow-up of children attending infant language units: Outcomes at 11 years of age. *International Journal of Language and Communication Disorders*, *36*, 207–219.
Cooper, D. H., Roth, F. P., Speece, D. L., & Schatschneider, C. (2002). The contribution of oral language skills to the development of phonological awareness. *Applied Psycholinguistics*, *23*, 399–416.
Curenton, S. M. (2004). The association between narratives and false belief in low-income preschoolers. *Early Education and Development*, *15*, 121–145.
Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade* Washington, DC: U.S. Department of Education, National Center for Education Statistics.
Dickinson, D. K., McCabe, A., Anastaspoulos, L., Peisner-Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology*, *95*(3), 465–481.
Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446.
Eisenberg, S., Fersko, T. M., & Lundgren, C. (2001). The use of MLU for identifying language impairment in preschool children: A review. *American Journal of Speech-Language Pathology*, *10*, 323–342.
Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
Farran, D. C., Lipsey, M. W., Watson, B., & Hurley, S. (2007). *Balance of content emphasis and child content engagement in an Early Reading First program*. Paper presented at the American Educational Research Association, Chicago, IL.
Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research*, *47*, 1301–1318.
Fraivillig, J. L., Murphy, L. A., & Fuson, K. C. (1999). Advancing children's mathematical thinking in Everyday Mathematics classrooms. *Journal for Research in Mathematics Education*, *30*, 148–170.
Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225–256). New York, NY: Information Age Publishing.
Fuson, K. C. (1988). *Children's counting and concepts of number*. New York, NY: Springer-Verlag.
Galván Carlan, V. (2000). *Development of an instrument to assess the use of developmentally appropriate practices in teaching mathematics in early childhood classrooms*. Doctoral dissertation, University of Houston.
Galván Carlan, V., & Copley, J. V. (2000). *Early childhood mathematics profile observational checklist*. Houston. TX: University of Houston.
Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology*, *49*, 363–383.
Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*, 345–362.
Gibson, E. J., & Levin, H. (1975). *The psychology of reading*. Cambridge, MA: The MIT Press.
Glasgow, C., & Cowley, J. (1994). *Renfrew Bus Story test (North American Edition)*. Centreville, DE: Centreville School.
Griffin, S. (2004). Building number sense with Number Worlds: A mathematics program for young children. *Early Childhood Research Quarterly*, *19*, 173–180.
Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Raschotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, *79*, 192–227.
Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group randomized trials in education. *Educational Evaluation and Policy Analysis*, *1*, 60–87.
Horizon Research Inc. (2001). *Classroom observation protocol*. Chapel Hill, NC: Author.
Invernizzi, M., Sullivan, A., Swank, L., & Meier, J. (2004). *PALS pre-K: Phonological awareness literacy screening for preschoolers* (2nd ed.). Charlottesville, VA: University Printing Services.
Klein, A., Starkey, P., Clements, D. H., Sarama, J., & Iyer, R. (2008). Effects of a prekindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, *1*, 155–178.
Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher math talk. *Developmental Psychology*, *42*, 59–69.
Krull, J. L., & MacKinnon, D. P. (1999). Multilevel modeling in group-based intervention studies. *Evaluation Review*, *23*, 418–444.
Lampert, M., & Cobb, P. (2003). Communication and language. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 237–249). Reston, VA: National Council of Teachers of Mathematics.
Landry, S. (2007). *MClass:CIRCLE*. New York, NY: Wireless Generation.
Lappan, G., & Schram, P. W. (1989). Communication and reasoning: Critical dimensions of sense making in mathematics. In P. R. Trafton, & A. P. Shulte (Eds.), *New directions for elementary school mathematics* (pp. 14–30). Reston, VA: National Council of Teaching of Mathematics.
LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*, 815–852.
Lee, J. S., & Ginsburg, H. P. (2007). Preschool teachers' beliefs about appropriate early literacy and mathematics education for low- and middle-socioeconomic status children. *Early Education & Development*, *18*(1), 111–143.
Magnuson, K. A., Meyers, M. K., Rathbun, A., & West, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal*, *41*, 115–157.
Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2008). Young children's motivational beliefs about learning science. *Early Childhood Research Quarterly*, *23*, 378–394.
McGill-Franzen, A. (2010). The National Early Literacy Panel Report: Summary, commentary, and reflections on policies and practices to improve children's early literacy. *Educational Researcher*, *39*(4), 275–278.
McGuinness, D., & Morley, C. (1991). Gender differences in the development of visuo-spatial ability in pre-school children. *Journal of Mental Imagery*, *15*, 143–150.
Molfese, V. J., Beswick, J. L., Molnar, A., & Jacobi-Vessels, J. (2006). Alphabetic skills in preschool: A preliminary study of letter naming and letter writing. *Developmental Neuropsychology*, *29*, 5–19.
Molfese, V. J., Modglin, A. A., Beswick, J. L., Neamon, J., Berg, S., & Berg, J. (2006). Letter knowledge, phonological processing, and print knowledge skill development in nonreading preschool children. *Journal of Learning Disabilities*, *39*(4), 296–305.
NAEYC. (1991). *Early childhood classroom observation—National Academy of Early Childhood Programs* (revised ed.). Washington, DC: National Association for the Education of Young Children.
National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Reston, VA: Author.
National Institute of Child Health and Human Developomental Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom

quality and teacher and student behavior. *The Elementary School Journal*, *105*(3), 305–323.

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

National Research Council. (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

National Research Council. (2009). *Mathematics in early childhood: Learning paths toward excellence and equity*. Washington, DC: National Academy Press.

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

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, *87*, 224–240.

O'Neill, D. K., Pearce, M. J., & Pick, J. L. (2004). Preschool children's narratives and performance on the Peabody Individualized Achievement Test – Revised: Evidence of a relation between early narrative and later mathematical ability. *First Language*, *24*(2), 149–183.

Pankratz, M. E., Plante, E., Vance, R., & Insalaco, D. M. (2007). The diagnostic and predictive validity of The Renfrew Bus Story. *Language, Speech, and Hearing Services in Schools*, *38*, 390–399.

Paul, R., Hernandez, R., Taylor, L., & Johnson, K. (1996). Narrative development in late talkers: Early school age. *Journal of Speech and Hearing Research*, *39*, 1295–1303.

Pearson, P. D., & Hiebert, E. H. (2010). National reports in literacy: Building a scientific base for policy and practice. *Educational Researcher*, *39*, 286–294.

Preschool Curriculum Evaluation Research Consortium. (2008). Effects of preschool curriculum programs on school readiness (NCER 2008-2009). Washington, DC: Government Printing Office.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2006). *HLM: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Razel, M., & Eylon, B.-S. (1986). Developing visual language skills: The Agam Program. *Journal of Visual Verbal Languaging*, *6*(1), 49–54.

Razel, M., & Eylon, B.-S. (1990). Development of visual cognition: Transfer effects of the Agam program. *Journal of Applied Developmental Psychology*, *11*, 459–485.

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

Sarama, J., & Clements, D. H. (2009b. April). *Scaling up successful interventions: Multidisciplinary perspectives*. Paper presented at the American Educational Research Association, San Diego, CA.

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

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

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

Smith, M. W., & Dickinson, D. K. (1994). Describing oral language opportunities and environments in Head Start and other preschool classrooms. *Early Childhood Research Quarterly*, *9*, 345–366.

Snow, C. E., Barnes, W. S., Chandler, J., Goodman, I. F., & Hemphill, L. (1991). *Unfulfilled expectations: Home and school influences on literacy*. Cambridge. MA: Harvard University Press.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities. *Journal of Educational Psychology*, *86*, 24–53.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language & Hearing Research*, *41*, 407–418.

Swank, P. R., Assel, M. A., Anthony, J. L., King, T., Gunnewig, S., & Landry, S. H. (2006). *Psychometric analyses of English m:Class data*. Paper presented at the annual meeting of the Society of the Scientific Study of Reading, Vancouver, Canada.

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

Treiman, R., & Bourassa, D. C. (2000). The development of spelling skill. *Topics in Language Disorders*, *20*(3), 1–18.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1994). Components of reading ability: Issues and problems in operationalizing word identification, phonological coding, and orthographic coding. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities* (pp. 279–333). Baltimore, MD: Brookes.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology*, *85*, 83–103.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing (CTOPP). Examiner's manual*. Austin, TX: PRO-ED.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement* (3rd ed.). Itasca, IL: Riverside.

Yore, L. D., & Hand, B. M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, *25*, 689–725.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychonomic Bulletin*, *131*, 3–29.