Effects of a multimedia enhanced reading buddies program in kindergarten and fourth grade

Article *in* The Journal of Educational Research · April 2017

CITATIONS
0

READS
60

7 authors, including:

Rebecca Deffes Silverman
University of Maryland, College Park
43 PUBLICATIONS **533** CITATIONS
SEE PROFILE

Young-Suk Kim
University of California, Irvine
68 PUBLICATIONS **982** CITATIONS
SEE PROFILE

Daniel McNeish
University of North Carolina at Chapel Hill
34 PUBLICATIONS **98** CITATIONS
SEE PROFILE

The Journal of Educational Research

ISSN: 0022-0671 (Print) 1940-0675 (Online) Journal homepage: http://www.tandfonline.com/loi/vjer20

Effects of a multimedia enhanced reading buddies program on kindergarten and Grade 4 vocabulary and comprehension

Rebecca Silverman, Young-Suk Kim, Anna Hartranft, Stephanie Nunn & Daniel McNeish

To cite this article: Rebecca Silverman, Young-Suk Kim, Anna Hartranft, Stephanie Nunn & Daniel McNeish (2016): Effects of a multimedia enhanced reading buddies program on kindergarten and Grade 4 vocabulary and comprehension, The Journal of Educational Research, DOI: 10.1080/00220671.2015.1103690

To link to this article: http://dx.doi.org/10.1080/00220671.2015.1103690

THE JOURNAL OF EDUCATIONAL RESEARCH
http://dx.doi.org/10.1080/00220671.2015.1103690

Routledge
Taylor & Francis Group

Effects of a multimedia enhanced reading buddies program on kindergarten and Grade 4 vocabulary and comprehension

Rebecca Silverman[a], Young-Suk Kim[b], Anna Hartranft[c], Stephanie Nunn[c], and Daniel McNeish[c]

[a]Department of Counseling, Higher Education, and Special Education, University of Maryland, College Park, Maryland, USA; [b]Florida Center for Reading Research, Department of Psychology, Florida State University, Tallahassee, Florida, USA; [c]Language and Literacy Research Center, University of Maryland, College Park, Maryland, USA

ABSTRACT

Reading buddies programs, which pair older and younger students to read books together on a regular basis, are common in many U.S. elementary schools. Yet, the research base on these programs is limited. Therefore, we conducted a quasiexperimental study of a reading buddies program targeting vocabulary and comprehension. The program we studied paired fourth-grade students with kindergarten students to read, talk, play, and write together. In all, 16 Grade 4 classrooms and 16 kindergarten classrooms participated in the treatment group and in the comparison group. The treatment included 10 one-hour sessions implemented over the course of roughly 10 weeks. Analyses revealed effects of treatment on proximal measures of vocabulary for both kindergarteners and fourth-grade students. However, there were no effects on distal measures for either group. Teachers' perceptions of the program are presented, and findings are discussed in light of the extant literature.

ARTICLE HISTORY
Received 11 June 2015
Revised 10 August 2015
Accepted 11 September 2015

KEYWORDS
Comprehension; cross-age peer learning; elementary; vocabulary

The Common Core State Standards, adopted by the majority of U.S. states and territories (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), emphasize outcomes for vocabulary and comprehension at all grade levels, but they do not give teachers guidance on how to support students' vocabulary and comprehension development (International Reading Association, 2012). Given that 65% of the nation's fourth-grade students are not proficient in reading (National Center for Education Statistics, 2013), the emphasis on vocabulary and comprehension seems warranted. However, to effectively address vocabulary and comprehension in schools, educators need more information about the efficacy of programs that target these skills.

Much of the attention to these vocabulary and comprehension will occur in schools' core reading programs. In addition, supplemental programs may serve to enhance students' vocabulary and comprehension, especially if these programs provide students with extra attention, interaction, and practice to further develop the vocabulary and comprehension skills students are acquiring (e.g., Apthorp et al., 2012; Fuchs & Fuchs, 2005; Goodson, Wolf, Bell, Turner, & Finney, 2011). A recent study in typical elementary schools found that 75% of classroom talk was attributed to teachers (Silverman et al., 2014). Assuming this is the norm, there is little time or space for students to talk and actively use words they have learned and skills they are developing. Including a peer-learning program as a supplement to regular instruction may bolster students' vocabulary and comprehension by providing students the time and space to talk with one another.

Reading Buddies programs are one type of peer-learning program that is implemented in many schools across the United States. In these programs, older and younger students are paired together to read and talk about books. Because research suggests that reading and talking about books can facilitate students' vocabulary and comprehension development (Kamil et al., 2008; National Reading Technical Assistance Center, 2010; Shanahan et al., 2010), it seems likely that reading buddies programs would be effective in supporting students in these areas. However, the research base on the efficacy of reading buddies programs is limited. Thus, to add to the research on reading buddies programs, we conducted a study to evaluate the effect of a particular reading buddies program on the vocabulary of kindergarten Little Buddies and the vocabulary and comprehension of fourth-grade Big Buddies. The program we studied included multimedia (i.e., video), reading and writing, and games and activities. Including multimedia in this program was meant to prime student learning and promote student engagement. In all, 16 Grade 4 classrooms and 16 kindergarten classrooms participated in the treatment group and 16 Grade 4 classrooms and 16 kindergarten classrooms participated in the comparison group. Before and after the intervention, we assessed all students in the study via a proximal measure of target vocabulary. Additionally, we assessed kindergarten students using a distal measure of general vocabulary knowledge and fourth-grade students via a distal measure of general comprehension skill. Finally, to determine how teachers felt about the program, we collected survey data on teachers' perceptions of the strengths and weaknesses of the program. Data on teachers' perceptions about the program was intended to provide an indication of

CONTACT Rebecca Silverman rdsilver@umd.edu Department of Counseling, Higher Education, and Special Education, University of Maryland, 1308 Benjamin Bldg, College Park, MD 20742.

the social validity of the program and to inform future development of cross-age peer learning programs. Subsequently we discuss the theoretical rationale for reading buddies programs and review related research. We also present details of the study we conducted and discuss the results in the context of the related research base.

Theoretical rationale

According to sociocultural theory (Vygotsky, 1978), children learn best through social interaction with more knowledgeable others who can support language development and model how to approach challenging tasks. While more knowledgeable others are often parents and teachers, peers can serve in this role as well. In cross-age peer learning programs, older children can serve as more knowledgeable others and provide guidance and support to younger children (Topping, 2005). The additional one-on-one attention and encouragement from older children may help motivate younger children (Slavin, 1996). Additionally, younger children can learn from the modeling and direction of the older children as well. Older children can also benefit from participating in a cross-age peer learning program by engaging in the metacognitive process of guiding a younger peer to understand new material (Gartner, Kohler, & Riessman, 1971; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003). Acting as role models for younger children can instill in older children a sense of responsibility and confidence (DePaulo et al., 1989; Schneider & Barone, 1997). By providing guidance and giving direction, older children may internalize and appropriate the content and strategies themselves (Cohen, Kulik, & Kulik, 1982).

Related research

Research over the past 20 years has yielded strong evidence of effective instruction to support vocabulary and comprehension. A recent review of the research on vocabulary instruction by the National Reading Technical Assistance Center (2010) suggested that "frequent exposure to targeted vocabulary," "explicit instruction of targeted vocabulary words," and "questioning and language engagement" are characteristics of effective vocabulary programs (p. 7). Additionally, Manyak et al. (2014) recommended four principles of vocabulary instruction for elementary students: "establish efficient yet rich routines for introducing target words; provide review experiences that promote deep processing of target words; respond directly to student confusion by using anchor experiences; and foster universal participation and accountability" (p. 16). Similarly, two recent reviews of research on comprehension instruction suggest that fostering high-quality discussion of text, supporting students' use of reading comprehension strategies (e.g., predicting, asking and answering questions, summarizing), and establishing motivating and engaging contexts for reading are important for developing comprehension skills (Kamil et al., 2008; Shanahan et al., 2010). Thus, any program aiming to support students' vocabulary and comprehension should include these aspects of instruction.

In the current digital age, incorporating multimedia, including videos and electronic texts in addition to traditional books, may be useful in supporting vocabulary and comprehension (Silverman & Hines, 2009; Dalton, Proctor, Uccelli, Mo, & Snow, 2011). Multimedia provides visual scaffolding and nonverbal supports, which can facilitate word learning and comprehension, particularly for students with limited vocabulary knowledge or comprehension skills (e.g., Kamil, Intrator, & Kim, 2000; Neuman, 1992). For example, using multimedia to support vocabulary and comprehension has been shown effective with English language learners (ELL) and their non-ELL peers in early and later elementary school (Silverman & Hines, 2009; Chambers et al., 2008; Proctor, Dalton, & Grisham, 2007).

In addition to the instructional components discussed thus far, peer learning, which has been found to be effective for supporting a range of reading skills including word recognition, fluency, self-concept, and motivation (Cohen et al., 1982; Puzio & Colby, 2013; Rohrbeck et al., 2003), has been explored as a context for vocabulary and comprehension development. There is a substantial body of research on the effects peer learning programs on comprehension. Programs such as Reciprocal Teaching (Lysynchuk, Pressley, & Vye, 1990; Palinscar & Brown, 1984; Rosenshine & Meister, 1994), Collaborative Strategic Reading (CSR; Klingner & Vaughn, 1998, 1999; Vaughn, Klingner, & Bryant, 2001), and Peer-Assisted Learning Strategies (PALS; e.g., Fuchs, Fuchs, & Burish, 2000; Fuchs, Fuchs, Mathes, & Simmons, 1997; Saenz, Fuchs, & Fuchs, 2005), have been implemented widely and shown to have positive effects on comprehension across studies. Reciprocal Teaching and CSR, in which heterogeneous groups of students work together to read and comprehend text, provide students with practice previewing text, monitoring comprehension and clarifying understanding, summarizing what was read, and asking and answering questions about text. As students help each other use these strategies, they become more proficient in using reading strategies on their own. The PALS program pairs more and less proficient readers to predict, read, and summarize together. Students are taught to support each other and give each other feedback in the program. Results from studies of Reciprocal Teaching, CSR, and PALS show positive effects for students in general education classrooms (Kelly, Moore, & Tuck, 1994; McMaster, Fuchs, & Fuchs, 2007) as well as those with learning disabilities (Klingner & Vaughn, 1996; Fuchs et al., 1997) and English language learners (ELLs; Klingner & Vaughn, 2000; Saenz et al., 2005).

While Reciprocal Teaching, CSR, and PALS employ a same-age peer learning model, other programs targeting comprehension have used cross-age peer learning models. For example, Van Keer and Vanderlinde (2010) investigated the effects of a cross-age peer learning program to promote reading strategy awareness, cognitive and metacognitive reading strategy use, and reading comprehension achievement with third- and sixth-grade students. Results from a quasiexperimental study showed significant effects of the intervention on third and sixth-grade students' awareness of reading strategies and reading strategy use. Additionally, Van Keer and Verhaeghe (2005) studied the impact of a cross-age peer learning program on second- and fifth-grade students reading comprehension and self-efficacy perceptions. Findings revealed that cross-age peer learning was supportive for both younger and older students.

Beyond comprehension, few studies have specifically investigated the role of peer learning in supporting vocabulary development. However, a couple of studies suggest that peer

learning may, indeed, be supportive of word learning. For example, Christ and Wang (2012) studied buddy reading in preschool classrooms and found that social interactions during buddy reading promoted students' use and learning of words. Additionally, Klingner and Vaughn (2000) found that diverse classrooms of same-age English proficient and ELL fifth-grade students actively discussed academic vocabulary with each other while participating in small group collaborative strategic reasoning. Furthermore, Zhang, Anderson, and Nguyen-Jahiel (2013) found that ELL students in Grade 5 used more diverse vocabulary in their writing after participating in a peer-led, open-format discussion approach called Collaborative Reasoning.

While these studies support using same-age peer learning to support vocabulary, other studies suggest that cross-age peer learning may also be effective for supporting word learning. For example, 7 and 11-year-old students who participated in a cross-age peer learning program targeting math skills showed increase use of math vocabulary across the five-week program (Topping, Campbell, Douglas, & Smith, 2003). Additionally, Topping, Peter, Stephen, and Whale (2004), studying cross-age peer learning focused on science with 7–8-year-olds and 8–9-year-olds, found that students who participated in the program made significant gains in understanding scientific concepts and keywords compared to their peers who did not. Neither of these programs focused specifically on vocabulary teaching and learning. Therefore, to determine the potential of cross-age peer learning programs to support vocabulary, research is needed on programs that focus specifically on vocabulary and that include components of effective vocabulary instruction outlined previously.

Several studies have explored the relative efficacy of same-age versus cross-age peer learning programs. For example, in the Van Keer and Verhaeghe (2005) study mentioned previously, researchers compared the effects of same-age and cross-age peer learning on reading comprehension for students in Grades 2 and 5. Results showed that cross-age peer learning benefitted second and fifth-grade students more than same-age peer learning. Topping, Thurston, McGavock, and Conlin (2012), working with a large sample of 8- and 10-year-old elementary school students, also compared the effects of same-age and cross-age peer learning on reading achievement. While same-age and cross-age peer learning showed similar effects in the short term (i.e., over the course of a one-year period), cross-age peer learning proved more beneficial to students over the long term (i.e., over a two-year period).

In summary, while the body of research on cross-age peer learning for vocabulary and comprehension provides some evidence of its effectiveness, there has been limited research on cross-age peer learning for vocabulary or comprehension with children in lower elementary school. Given that many schools implement reading buddies programs that partner older and younger students to read together (e.g., Babicki & Luke, 2007; Davenport, Arnold, Lassmann, & Lasmann, 2004; Lowery, Sabis-Burns, & Anderson-Brown, 2008) and given that little research exists on the efficacy of these programs for supporting the language and literacy development of older and younger children, research on cross-age peer learning programs with students in lower and upper elementary school is needed.

A final element to consider when evaluating a new program to promote vocabulary and comprehension is its social validity (e.g., Leko, 2014; Lindo & Elleman, 2010). How teachers who implement the program view its feasibility and efficacy is an important indicator of the practicality and sustainability of the program. Specifically, if teachers feel that a program is difficult to implement or does not have a positive effect on their students, they are unlikely to devote the time and energy needed to adopt the program and implement it with fidelity. If teachers do not have faith in the program, then the work spent developing materials and training teachers to implement the program is wasted. Thus, understanding teachers' perceptions of a program, including whether the program seems easy to implement and beneficial for students, is worth studying when evaluating new programs. Furthermore, teacher perceptions of a program can guide future development and revision of programs to support student learning.

Present study

The reading buddies program we evaluated is called the *Martha Speaks Reading Buddies* (MSRB) program (WGBH Educational Foundation, 2008). The program was developed by WGBH Boston, producers of the children's television program called *Martha Speaks* (WGBH Boston, 2008). This program is sponsored by the Public Broadcasting Service (PBS) and the Corporation for Public Broadcasting (CPB), and the program is based on a children's book by Susan Meddaugh entitled *Martha Speaks* (Meddaugh, 1995). The program and the book feature a dog, Martha, who learns to talk. The educational goal of the program is to foster vocabulary. Each episode of the *Martha Speaks* program focuses on a specific theme and includes two 11-min stories. In each 11-min story, 4–5 target words are explicitly defined and repeated roughly five times. Staff from WGBH Boston's educational outreach office created the MSRB program to foster vocabulary in the school setting. The program, which can be accessed at http://www.pbs.org/parents/martha/readingbuddies/index.html, pairs fourth-grade Big Buddies and kindergarten Little Buddies. Together, reading buddies (a) watch an 11-min *Martha Speaks* story, (b) talk about target words from the show, (c) play a game or do an activity that focuses on the target words, (d) read a book that relates to the target words, and (e) write or draw about something related to the target words in a special journal. At the time the program was implemented for this evaluation, it consisted of 10 reading buddies sessions.

The present study was commissioned and funded by PBS, CPB, and a state public media association. Dr. Rebecca Silverman served as a consultant to WGBH Boston and PBS and CPB during the time of the study. However, the researchers on the study, including the lead author, designed the evaluation, oversaw the data collection, and analyzed the data independently and without bias. To conduct the evaluation, we worked with local PBS affiliates to recruit teachers and students. They assigned teachers and their students to a treatment or comparison group. While the local PBS affiliates trained the teachers, we independently trained external research assistants who were not informed of the intent of the research to assess students in the treatment and comparison groups before and after program implementation. We oversaw the assessment process and handled all data analyses ourselves. We used hierarchical linear modeling to answer the following research questions:

Research Question 1: What are the effects of the MSRB program on kindergarten students' program-specific and general word knowledge? Do these effects differ for children with higher and lower vocabulary at the start of the program?

Research Question 2: What are the effects of the MSRB program on fourth-grade students' program-specific word knowledge and general comprehension skill? Do these effects differ for children with higher and lower vocabulary at the start of the program?

In addition to the effects of the program, we were also interested in teachers' perceptions about the strengths and weaknesses of the program that could inform future development of cross-age peer learning programs. Thus, we used qualitative data analysis to answer the following additional research question:

Research Question 3: According to teachers who participated in the MSRB program, was the program effective and feasible? What were the strengths and weaknesses of the program?

Methods

Sample

The sample was drawn from 16 schools across 16 school districts located within the same state in the southeast region of the United States. The local PBS affiliates asked teachers in these schools to volunteer for participation in the study. In each school, at least two kindergarten teachers and two fourth teachers volunteered and the principals chose which teachers would participate in the study. Principals were asked to choose effective teachers. Within these schools, teachers and their classes were randomly assigned to the treatment or comparison group. For program implementation, the kindergarten class assigned to the treatment group was paired with the Grade 4 class that was assigned to that group as well. Thus, there were 32 teachers in the treatment group and 32 teachers in the comparison group. All participating teachers in the treatment group agreed to implement the program with their whole class. Approval for program implementation was obtained from each district represented in the study. All participating teachers in the comparison group agreed to implement instruction as usual via the typical curriculum and instructional materials for their grade level, but they were promised that all program materials would be provided to them after the program was completed in case they wanted to implement the program at another time.

After teacher recruitment, we recruited children to participate in the evaluation of the program by sending home permission forms to parents of all children in participating teachers' classrooms. These permission forms asked parents to allow us to assess their children before and after program implementation. Parents were not informed beforehand whether their children were in the treatment or comparison group. They were told that their children may or may not participate in the program depending on whether their classroom teacher was randomly assigned to the treatment or comparison group. While all children in the classrooms of teachers in the treatment group participated in the program, we only assessed children for whom we had parental permission. We obtained parental permission for roughly two-thirds of the children in participating teachers' classrooms to participate in the study. In all, 447 kindergarten children and 479 fourth-grade children participated in the evaluation of the program. In kindergarten, 236 children were in the treatment group and 211 were in the comparison group. In Grade 4, 265 children were in the treatment group and 214 children were in the comparison group. Table 1 shows characteristics of the student sample by grade level and group.

Program

Materials

Teachers in the treatment group were provided with a kit of materials that included a teachers' guide, which described the theoretical rationale for the program, instructions for setting up the program and creating buddy pairs within their classrooms, and an implementation checklist. Teachers were also given student materials, which included a guide for the Big Buddies to follow when working with their Little Buddies. Finally, teachers were provided with books for children to read together, DVDs for children to watch together, and pencils and stickers for children to take home. The instructional content was designed around a *Martha Speaks* video and a related text. The *Martha Speaks* video contained four target vocabulary words at the Tier 2 level (see Beck, McKeown, & Kucan 2002) that were explicitly defined at least once and repeated several times throughout the episode. The related books for each video and lesson were selected based on their thematic relation to the episode, reading level, and interest for children. The thematic link between the videos and texts fostered students' semantic network of the content and encouraged students to use the target vocabulary in multiple activities throughout the lesson. The reading level for all of the texts was set to below Grade 4 so that the Big Buddies would be able to read with fluency and understanding.

Structure

Prior to beginning the program, fourth-grade and kindergarten teachers met with researchers to review program components and receive training on the materials found in their MSRB kit. Program materials addressed particulars such as planning, scheduling, and pairing. For example, program materials suggested that kindergarten and fourth-

Table 1. Demographic information for students in the sample.

	Kindergarten		Grade 4	
	Treatment ($n = 264$)	Comparison ($n = 240$)	Treatment ($n = 290$)	Comparison ($n = 244$)
White	64%	59%	65%	61%
Black	30%	33%	28%	31%
Other	6%	8%	7%	8%
NSLP eligible	61%	64%	58%	62%
English learner	14%	10%	4%	5%
Female	48%	54%	46%	52%

Note. NSLP = National School Lunch Program.

grade teachers should (a) meet to discuss upcoming reading buddies sessions together, (b) schedule reading buddies sessions before or after a transition time to facilitate moving kindergarteners to fourth-grade classrooms or vice versa, and (c) consider personalities (pair shy students with nurturing students), backgrounds (pair English learners who share home language) and ability levels (pair higher readers together and lower readers together so pairs will be roughly evenly matched and provide extra support to pairs that might struggle) when pairing students. At the start of the program, teachers led an introductory session to teach students about the roles of the Big and Little Buddies. Teachers reviewed the Big Buddy guide with the fourth-grade students during the introductory session. Following the introductory session, the program also included 10 sessions in which the buddies met together. Each session, big and little buddies sat together to watch an 11-min *Martha Speaks* video. Then, they discussed questions about the video and played games related to the video. The discussion questions and related games featured and encouraged students to use four target words from the episode. These words had been defined and repeated roughly five times in the episode, and the definitions for these words were provided in the Big Buddy guide. The target words were the same for fourth-grade students and kindergarteners in order to allow them to learn collaboratively. As word knowledge develops along a continuum from no knowledge to some knowledge to deep knowledge (Beck, McKeown, & Kucan, 2002), words were chosen to be appropriate for both younger and older students such that kindergarteners could learn the words at an initial level and fourth-grade students could learn the words at a more advanced level. The games varied across sessions. For example, one game, called *Choose and Chat*, included question cards (e.g., "What are Martha's special *talents*?) and buddies were encouraged to use target words in their responses to one another. In another game, called *Tic-Tac-Talk*, buddies were given game boards with pictures on them. Buddies took turns marking squares, as in tic-Tac-Toe, and were encouraged to use target words to describe the picture (e.g., using *real* vs. *pretend*). During one session, buddies played a game called *Skits' Tricks* at the *Martha Speaks* website on the computer. Each session, after playing the game, Big Buddies read a book aloud to the Little Buddies. This book was related to the theme of the show and target words the buddies had already encountered in the session. For example, after watching the *Martha Speaks* story called *Martha and Skits* that focused on the *unique* talents of the Martha and Skits characters, talking about the word *unique* and what it means, and playing a game that featured the word *unique*, Big Buddies read the book *Star of the Week* by Barny Saltzberg (2010) to their Little Buddies. This book is about a character named Stanley who shares all of the unique things about himself with his class as the star of the week. After reading the book, the Big Buddy read a question and the buddies drew and wrote in response to the question. For example, in a reading buddies session focused on the word *courageous*, buddies drew a picture and wrote a journal entry about what courageous thing they would do as a superhero.

Training

Teachers participating in the treatment group were trained in person by staff from the PBS affiliates that recruited schools and teachers. The trainers had been trained by WGBH staff at an all-day in-person training before the start of the study. Training for teachers consisted of a review of program materials and a discussion of program implementation. Teachers were told that their role during the buddy sessions was to monitor students and facilitate buddy interaction when needed. Training at each school lasted approximately 2 hr.

Fidelity

In order to examine the extent to which the MSRB program was implemented as originally intended, research assistants at each location observed program implementation two times. Observers completed a checklist of program elements that were to be implemented. The following is a list of some of the elements that observers tracked: (a) *Martha Speaks* episode is played, (b) buddies talk about the episode, (c) buddies play a game together, (d) buddies read a book together, (e) buddies talk about the book together, and (f) buddies write together. Observers tracked two pairs of students for each buddies element. On average, 90% of the primary elements were observed during buddy sessions.

Assessments

As schools in the project were concerned with overtesting students, we tried to limit the amount of testing we conducted at each grade level to 45 min per student. We reasoned that if the intervention was effective the effect would be most evident on proximal measures (e.g., program-specific word learning in both kindergarten and Grade 4) so we gave measures of target vocabulary knowledge in each grade. The target vocabulary measures took roughly 30 min per student to administer. Thus, we had only 15 min of testing left for more distal (i.e., generalizable) measures. While we would have preferred to assess generalizable vocabulary and comprehension in both kindergarten and Grade 4, we did not have enough time to administer assessments of both outcomes at both grade levels. Since most kindergarteners are not yet reading, it is more appropriate to measure kindergarten language comprehension than reading comprehension (Catts, Adlof, & Weismer, 2006). Vocabulary is a strong proxy for language comprehension and it is predictive of later reading comprehension (e.g., Cunningham & Stanovich, 1997); thus, we decided to measure generalizable vocabulary in kindergarten. However, as it is appropriate to assess fourth-grade students on generalizable reading comprehension and since the primary goal of reading (i.e., the *sine qua non* of reading) is reading comprehension (Beck & McKeown, 2007), we decided investigate whether the intervention had effects on generalizable comprehension in Grade 4. Future studies should include generalizable measures of both vocabulary and comprehension at each grade level.

Kindergarten target word knowledge

The Test of Word Knowledge–Kindergarten (TWK-K) included two subtests: (a) a picture subtest and (b) a definition subtest. In the picture task, students were asked to choose

which picture of four matched the word given by the administrator. Every word targeted in the MSRB program was assessed on the picture subtest. Thus, there were 40 items on the picture subtest. In the definition task, children were asked to tell what each word means. The rubric for scoring the definition task was on a scale of 0 to 2. Students received a score of 2 if they provided a clear and concise definition of the target word. Students received a score of 1 if they provided an example of the target word or concepts related to the target word. Students received a score of 0 if they provided an unrelated example or incorrect use of the target word. Only half of the words targeted in the MSRB program were assessed on this subtest. Thus, there were 20 items on this subtest. The reliability (Cronbach's alpha) of the picture subtest was .79 at pretest and .82 at posttest. The reliability (Cronbach's alpha) of the definition subtest was .69 at pretest and .80 at posttest. Interrater reliability (Cohen's kappa) of the definition task ranged from .82 to .94. Note that standard practice assumes that, in general, greater than or equal to .70 is acceptable and greater than or equal to .60 is acceptable for newly developed measures (Gersten et al., 2005).

Kindergarten general vocabulary knowledge

The Peabody Picture Vocabulary Test IV (PPVT; Dunn & Dunn, 2007) was used to assess children's general vocabulary knowledge. Form A was used at pre- and posttest. This is a commonly used norm-referenced measure of receptive vocabulary in which children choose one out of four pictures that corresponds to the target word given orally by the test administrator. This assessment takes 15–20 min to administer. This receptive measure of vocabulary knowledge is norm-referenced. Standard scores ($M = 100$, $SD = 15$) were used in analyses. The PPVT is correlated with the Clinical Evaluation of Language Fundamentals-4 Core Language Composite (.73 for ages 5–8), the Expressive Vocabulary Test-2 (average correlation of .82), and the Group Reading Assessment and Diagnostic Evaluation Total Test Score (.58). Cronbach's alpha is reported to range from .95 to .97 for kindergarten-aged children (Dunn & Dunn, 2007).

Grade 4 target word knowledge

The Test of Word Knowledge-Fourth Grade (TWK-4) administered to big buddies at pretest and posttest consisted of multiple-choice questions asking students to choose the word that best completed given sentences. For example, one question asked, "Someone who does not get scared when there is danger is _____. fearful, real, brave, or honest." Every word targeted in the program was assessed on this measure. Thus, there were 40 items on the measure. Reliability (Cronbach's alpha) estimates were .82 at pretest and .84 at posttest.

Grade 4 comprehension

The Test of Sentence Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010) was a group-administered reading comprehension assessment given at pre- and posttest to fourth-grade students in the study. In this assessment students were given three min to read and respond true or false to a series of single sentence items (e.g., A doughnut is made of very hard steel). This measure has been shown to have high correlations with the Gates MacGinitie Reading Test Comprehension Subtest and the Woodcock Munoz Language Survey Passage Comprehension Subtest and a latent variable of reading comprehension composed of multiple measures (Leung, Silverman, Nadakumar, & Hines, 2011; Proctor, Silverman, Harring, & Montecillo, 2012).

Additionally, the Dynamic Indicators of Basic Literacy Skills DAZE task was administered in groups to all fourth-grade students. In the DAZE task, student silently read a passage. The there is a blank and three choices at every seventh word in the passage. Students are expected to choose the word that fits correctly in the blank. Students are given three min to correctly complete as many items as possible. The DAZE task has a moderately strong relation with the group reading assessment and diagnostic evaluation total score with correlation coefficients ranging from .63 to .68 for students in Grade 4 (Good et al., 2010).

Teacher survey

At the completion of the program, kindergarten and fourth-grade teachers who implemented the program completed a brief survey about their perceptions of the program. For the purposes of this study, we analyzed responses to the following prompts or questions: (a) Please rate how difficult the *Martha Speaks* program was to implement: easy, fairly easy, moderate, hard, very hard. (b) Were there any barriers to program implementation? Please check yes or no and please explain. (c) Did the program fit easily into your classroom lesson plans? Please check yes or no and please explain. (d) Do you feel that your students benefited from participating in the program? Please check yes or no and please explain. (e) What were the strengths of the program? (f) Did the program help with any of the following? Check all that apply: vocabulary, comprehension, fluency, motivation, engagement. (g) What suggestions do you have for improving the program? The research team distributed the survey to classroom teachers in paper and electronic formats. Only 75% of teachers who implemented the MSRB program responded to the survey.

Analyses

Student outcomes

Because participants responded to multiple measures that tapped different, but related constructs, the statistical analysis modeled all outcomes with a single, multivariate model for each grade separately to preserve power to detect intervention effects. A multivariate model allowed the covariance between the different measures to be taken into account, which ultimately reduces the amount of residual variance, increasing statistical power compared to modeling each individual outcome separately while also yielding regression coefficients that account for the dependency between the measures. Furthermore, this allowed estimates of effects that would be constant across multiple univariate models to be pooled if they were approximately equal, which increases degrees of freedom and, thus, augments power.

Intraclass correlations (ICCs) for students nested within teachers were fairly high (range = .25−.34), indicating that the clustering of students within classrooms was

non-ignorable. Therefore, the statistical model was a three-level mixed-effects model in which Level 1 consisted of the posttest scores of the measures taken by students, Level 2 comprised students and student-level predictors (pretest scores, free or reduced lunch status), and Level 3 was the classroom level and classroom level predictors (treatment group status). The covariance structure of the residuals at Level 1 was modeled as unstructured, meaning that each element of the lower diagonal was freely estimated to account for the relation between the different outcome measures. Although the unstructured approach is the least parsimonious, it was selected because the residual variances were quite different between the different outcomes and the covariances did not appear to follow a more parsimonious pattern (e.g., Toeplitz or compound symmetric). A random effect was included for the outcome-specific intercepts at Level 3 but random effects were not included at Level 2 because the model without Level 2 random effects was more parsimonious and fit just as well as a model with a Level 2 intercept based on a likelihood ratio test (kindergarten: $\chi^2_1 < 0.10$, $p > .75$; Grade 4: $\chi^2_1 = 0.10$, $p = .75$).

As the number of classrooms was fairly small at 32, we used Kenward-Roger degrees of freedom and covariance matrix adjustments (Kenward & Roger, 1997, 2009), which have been shown to provide unbiased coefficients and standard error estimates with as few as ten clusters (e.g., Bell, Morgan, Schoeneberger, Kromrey, & Ferron, 2014; McNeish & Stapleton, 2014). The Kenward-Roger correction allows for different degrees of freedom for each individual estimate and fractional degrees of freedom; thus, there are different degrees of freedom represented in the subsequent results.

Teacher survey

We analyzed descriptive statistics from teachers' responses to yes–no and multiple-choice prompts and questions. We also coded teachers' open-ended comments for themes related to strengths and weaknesses of the program. To do so, two independent researchers jointly read through all comments and created a list of thematic codes that characterized the main gist of teacher comments within each open-ended item. These codes were created through the constant comparative method (Glaser & Strauss, 1967) and are listed in the Appendix. Once these codes were established, each researcher independently coded all comments from each question using this coding list. At the conclusion of coding, the researchers compared codes in order to determine interrater reliability, which was 100%.

Results

Descriptive statistics are presented in Table 2. Standard scores in the normed measures (i.e., PPVT and TOSREC) indicate that the mean scores of participating kindergartners' vocabulary and fourth-grade students' reading comprehension were in the average range across the treatment and comparison conditions. Results for the multivariate model for kindergarten and fourth-grade students are reported in Tables 3 and 4, respectively. Rather than report results as relative differences from a reference outcome as is output by statistical software, for ease of interpretation, we report absolute effects so that regression coefficient estimates can be interpreted as if three separate univariate models were run. Inferential tests in Tables 3 and 4 assess whether the effects are significantly different from zero and were obtained using Estimate statements in SAS Proc Mixed Version 9.3 (SAS Institute, Cary, NC).

Table 2. Means and standard deviations (raw scores unless otherwise indicated) by condition.

	Treatment				Comparison			
	Pretest		Posttest		Pretest		Posttest	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Kindergarten								
TWK-K picture	20.87	6.09	24.89	7.26	21.75	5.69	23.75	6.41
TWK-K definition	3.12	3.74	5.96	6.00	3.24	3.25	4.19	4.45
PPVT	98.28	17.96	98.99	15.53	100.22	15.95	101.48	16.12
Grade 4								
TWT-4	32.58	5.95	35.42	7.81	31.71	6.64	32.07	7.69
TOSREC	98.14	18.48	99.91	16.13	96.36	16.75	97.99	16.58

Note. Standard scores are reported for Peabody Picture Vocabulary Test IV (PPVT) and Test of Sentence Reading Efficiency and Comprehension (TOSREC). TWT-4 = Test of Word Knowledge–Grade 4; TWT-K = Test of Word Knowledge–Kindergarten.

Kindergarten results

Equity of treatment and comparison groups

To examine whether the demographic makeup of the treatment and comparison groups were relatively equal, we ran a logistic regression using group status as the binary outcome and sex, ethnicity (White, Black, other), ELL status, and socioeconomic status as implied by whether students were eligible to receive free or reduced lunch via the National School Lunch Program (NSLP). The clustering of students within teachers was accounted for by using generalized estimating equations with an independent working correlation matrix and an MBN small sample correction that has been found to perform well with fewer than 40 clusters with binary outcomes (Morel, Bokossa, & Neerchal, 2003) because the population-average effects were of interest. The results indicated no differences between group assignment for sex, $F(1,31) = 0.93$, $p = .34$; ethnicity, $F(2, 41) = 0.21$, $p = .81$; ELL status, $F(1, 422) = 0.28$, $p = .59$; or NSLP, $F(1, 422) = 0.20$, $p = .66$). Given research suggesting differences in vocabulary by NSLP (Fernald, Marchman, & Weisleder, 2013), however, we decided to keep NSLP as a control variable for further analyses.

Pooling predictors across outcomes

Control variables such as NSLP, pretest score, and the pretest by treatment interaction were tested with multiparameter Wald type III F tests and likelihood ratio chi-square tests (LRTs) to discern whether pooled estimates were more parsimonious than individual estimates of the same effect for each outcome. For the kindergarten model, Wald tests and LRTs suggest that pretest effects, $F(2,591) = 8.61$, $p = .01$, and NSLP, $F(2, 433) = 3.10$, $p = .05$, should be estimated separately for each outcome. The pretest by treatment interaction was reasonable to pool across outcome variables, $F(2, 433) = 0.94$, $p = .39$, meaning that the pretest by treatment interaction estimate was approximately equal across all three outcomes.

Covariance parameters

The intercepts were allowed to vary randomly at Level 3 to capture variability present at the classroom level. The variance of the intercept random effect at Level 3 was 1.79, ${}_{50:50}\chi^2(1, 2, N = 32) = 48.50$, $p < .01$, indicating that there is significant variability with regard to students average posttest score between classrooms on the outcome measures. Based on the model, residuals were meaningfully related and residual variances, residuals covariances, and their significance tests are reported in Table 3. We determined p values for variances with 50:50 mixture chi-square tests (Stram & Lee, 1994).

TWK-K—Picture subtest

No significant difference was observed between pretest measures and treatment status, $t(696) = 0.19$, $p = .85$. The treatment had a small effect based on Hedge's g and were predicted to score 1.70 points higher on the TWK-K picture subtest, $t(78) = 2.61$, $p = .01$, $g = 0.25$, holding all other predictors in the model constant. Students who qualify for the NSLP were also predicted to score 2.89 points lower on the TWK-K picture subtest, $t(412) = -5.19$, $p < .01$, $g = 0.59$, than students who did not qualify for the NSLP, holding all other predictors in the model constant.

TWK-K—Definition subtest

No significant difference was observed between pretest measures and treatment status, $t(696) = 0.19$, $p = .85$. The treatment had a borderline medium effect based on Hedge's g and were predicted to score 1.66 points higher on the TWK-K definition subtest, $t(48.1) = 2.92$, $p < .01$, $g = 0.28$, holding all other predictors in the model constant. Students who qualify for the NSLP were also predicted to score 2.41 points lower on the TWK-K definition subtest, $t(435) = -5.78$, $p < .01$, $g = 0.65$, than students who did not qualify for the NSLP, holding all other predictors in the model constant.

Table 3. Multivariate model estimates for kindergarten students.

Effect	Outcome	Estimate
Fixed effects		
Intercept		
	TWK-K DEF	5.57
	TWK-K PIC	24.88
	PPVT	103.91
Pretest		
	TWK-K DEF	0.86***
	TWK-K PIC	0.62***
	PPVT	0.65***
Treatment		
	TWK-K DEF	1.66**
	TWK-K PIC	1.70**
	PPVT	−0.69
Pretest by treatment		0.01
NSLP status		
	TWK-K DEF	−2.41***
	TWK-K PIC	−2.89***
	PPVT	−5.18***
Variance components		
Level 3 intercept var		1.79***
Level 1 var	TWK-K DEF	11.33***
Level 1 var	TWK-K PIC	21.56***
Level 1 var	PPVT	88.16***
Level 1 cov	TWK-K DEF, TWK-K PIC	6.11**
Level 1 cov	TWK-K DEF, PPVT	6.45***
Level 1 cov	TWK-K PIC, PPVT	5.62

Note: TWK-K = Test of Word Knowledge–Kindergarten; DEF = TWK-K definition subtest; PIC = TWK-K picture subtest; PPVT = Peabody Picture Vocabulary Test IV; NSLP = National School Lunch Program. We determined p values for variances by 50:50 mixture chi-square tests and p values for covariances were determined by Z tests.
*$p < .05$; **$p < .01$; ***$p < .001$.

PPVT

No significant difference was observed between pretest measures and treatment status, $t(696) = 0.19$, $p = .85$. Students in the treatment group performed no differently on the PPVT, $t(304) = -0.67$, $p = .51$, $g = 0.06$, than students in the comparison group, holding all other predictors in the model constant. Students who qualify for the NSLP were also predicted to score 5.18 points lower on the PPVT, $t(427) = -4.58$, $p < .01$, $g = 0.52$, than students who did not qualify for the NSLP, holding all other predictors in the model constant.

Grade 4 results

Equity of treatment and comparison groups

As we did with the kindergarten sample, we ran a logistic regression using group status the binary outcome and sex, ethnicity (White, Black, other), ELL status, and NSLP status. The results indicated no differences between group assignment for sex, $F(1, 31) = 0.39$, $p = .54$; ethnicity, $F(2, 34) = 0.15$, $p = .86$; ELL status, $F(1, 440) = 0.13$, $p = .72$; or NSLP status, $F(1, 440) = 0.49$, $p = .48$. As in the kindergarten analysis, we decided to keep NSLP as a control variable for further analyses.

Pooling predictors across outcomes

Similar to the kindergarten model, multiparameter Wald type III tests and LRTs were implemented to determine whether

Table 4. Multivariate model estimates for fourth-grade students

Effect	Measure	Estimate
Fixed effects		
Intercept		
	DAZE	29.08
	TOSREC	29.98
	TWK-4	33.53
Pretest		0.66***
treatment		
	DAZE	0.42
	TOSREC	0.59
	TWK-4	2.88***
Pretest by treatment		0.03
NSLP status		−0.70*
Variance components		
Level 3 intercept var		3.70***
Level 1 var	DAZE	37.37***
Level 1 var	TOSREC	40.52***
Level 1 var	TWK-4	24.38***
Level 1 cov	DAZE, TOSREC	10.17*
Level 1 cov	DAZE, TWK-4	4.26**
Level 1 cov	TOSREC, TWK-4	1.73

Note: DAZE = Dynamic Indicators of Basic Literacy Skills DAZE task; NSLP = National School Lunch Program; TOSREC = Test of Sentence Reading Efficiency and Comprehension; TWK-4 = Test of Word Knowledge - Grade 4. We determined p values for variances by 50:50 mixture chi-square tests and p values for covariances were determined by Z tests.
*$p < .05$; **$p < .01$; ***$p < .001$.

predictors could be pooled across the outcome measures. Both the Wald test, $F(3, 557) = 1.40$, $p = .25$; and LRT, $_{50:50(1,2)\ 2}\ \chi^2(2, N = 534) = 2.70$, $p = .26$, suggested the pretest by treatment interaction effect did not need to be estimated separately for each outcome and could be pooled. A Wald test, $F(2, 349) = 1.68$, $p = .19$; and LRT, $_{50:50(1,2)\ 2}\ \chi^2(2, N = 534) = 3.30$, $p = .19$, similarly suggested that the effect of NSLP status did not differ across outcomes and could be pooled. The effect for pretest was significantly different across outcomes based on a Wald test, $F(2, 562) = 3.29$, $p = .04$; and LRT, $_{50:50(1,2)\ 2}\ \chi^2(2, N = 534) = 7.70$, $p = .02$.

Covariance parameters

The intercept for each outcome was allowed to vary randomly at Level 3 to capture variability present at the classroom level. The variance of the random effect for intercept at Level 3 was 3.70, $_{50:50(1,2)\ 2}\ \chi^2(1, 2, N = 32) = 40.00$, $p < .01$, indicating that there is significant variability with regard to students' average posttest score between classrooms on the outcomes measures. Based on the model, residuals were meaningfully related and residual variances, residuals covariances, and their significance tests are reported in Table 4. We determined p values for variances with 50:50 mixture chi-square tests (Stram & Lee, 1994).

TWK-4

No significant difference was observed between pretest measures and treatment status, $t(863) = 0.68$, $p = .50$. The treatment group had a borderline small effect based on Hedge's g and were predicted to score 2.88 points higher on the TWK-4, $t(304) = 3.48$, $p < .01$, $g = 0.32$, than students in the comparison group, holding all other predictors in the model constant. Students who qualify for the NSLP were also not predicted to score significantly differently on the TWK-4, $t(303) = -1.67$, $p = .10$, $g = 0.19$, than students who did not qualify for the NSLP.

TOSREC

No significant difference was observed between pretest measures and treatment status, $t(863) = 0.68$, $p = .50$. Students in the treatment group did not score significantly different on the TOSREC from students in the comparison group, $t(267) = 0.67$, $p = .52$, $g = 0.06$, holding all other predictors in the model constant. Students who qualify for the NSLP did not score significantly different on the TOSREC from students who did not qualify for the NSLP, $t(303) = -1.67$, $p = .10$, $g = 0.19$.

DAZE

No significant difference was observed between pretest measures and treatment status, $t(863) = 0.68$, $p = .50$. Students in the treatment group did not score significantly different on the DAZE than students in the comparison group, $t(82.2) = 0.47$, $p = .64$, $g = 0.04$, holding all other predictors in the model constant. Students who qualify for the NSLP did not score significantly different on the DAZE from students who did not qualify for the NSLP, $t(303) = -1.67$, $p = .10$, $g = 0.19$.

Teacher survey results

In response to how difficult the program was to implement, 92% ($n = 22$) of the teachers responded that the program was easy or fairly easy to implement. Nearly 60% ($n = 14$) of the teachers reported that they experienced barriers to implementation. Coding the open-ended comments from those teachers, three themes emerged: finding time in their schedule to implement, pacing within the lessons, and use of technology. Teachers found it difficult to coordinate with their partners to find a time to execute the lessons together ($n = 5$). In addition, the lessons contained a lot of material, and teachers indicated that one challenge to implementing the program was completing all of the lesson activities in the time allotted ($n = 6$). Finally, one of the teachers indicated that using technology in the program was a barrier to implementation in her classroom.

When asked whether the program aligned with the curriculum standards in their classroom or school, 95.2% of the teachers felt that the program aligned well with the curriculum standards. When asked whether they thought students benefitted from participating in the program, all teachers responded that they felt the program benefitted students. They explained that the students learned words, learned to work cooperatively, and had fun. Teachers reported that strengths of the program included cross-age learning, use of multiple modalities, enriched vocabulary, and alignment with standards. Pooling across teacher responses, 71% ($n = 17$) of the teachers indicated that they felt the cross-age structure of the program was a strength for their students. The use of multiple modalities to present and reinforce learning was a strength mentioned by nearly 60% ($n = 14$) of the teachers. All of the teachers responded that the enriched new vocabulary that the program presented to students was a strength of the program. Finally, three of the teachers indicated that the program's alignment with curriculum standards was a strength of the program.

When asked to indicate which of the following literacy and socioemotional skills that they felt the program addressed vocabulary, comprehension, fluency, motivation, or engagement, 100% indicated that the program helped with their students' vocabulary skills, 83.3% indicated that the program helped with their students' comprehension skills, 72.7% indicated that the program helped with their students' fluency skills, 95.8% indicated that the program improved their students' motivation, and 87% indicated that the program improved their students' engagement.

Discussion

The cross-age peer learning program studied within this research project, namely, the MSRB program, was founded on many principles of evidence-based vocabulary and comprehension instruction (e.g., Silverman & Hines, 2009; Dalton et al., 2011; Kamil et al., 2008; Manyak et al., 2014; National Reading Technical Assistance Center, 2010; Shanahan et al., 2010; Topping et al., 2003; Topping et al., 2004; Van Keer & Vanderlinde, 2010; Van Keer & Verhaeghe, 2005). Specifically, the program included explicit definitions of vocabulary words with multiple opportunities for students to use and apply the words throughout each lesson. The program incorporated

also multimedia via videos and traditional books, which enabled student to use words across different contexts. Also, the multimedia in the program provided visuals and animation to support word learning. Although comprehension strategies were not directly taught within the program, buddies discussed the videos they watched and the texts they read as they played games and participated in collaborative activities. The cross-age buddies concluded each lesson with a writing activity that reinforced the target vocabulary.

Results from the present quasiexperimental study of the effectiveness of the program suggest that the program holds promise for supporting older and younger students' vocabulary learning. Specifically, as seen in analyses of proximal measures of student learning, the program had a positive effect on kindergarten and fourth-grade students' knowledge of words targeted in the program. And, these effects did not differ by whether children had lower or higher vocabulary knowledge at the beginning of the program. Findings are in line with previous research on cross-age peer learning that shows positive effects on vocabulary (e.g., Topping et al., 2003; Topping et al., 2004). Talking about words with a peer can be a useful support for vocabulary learning. Typically, in regular teacher-led vocabulary instruction, a few students may be given a chance to use a word and receive feedback on a given day. However, in a peer-based vocabulary program, all students get a chance to use and receive feedback on use of all targeted words. The cross-age context may have benefitted younger learners, who likely received more support for word learning than they would have from a same-age peer, as well as older learners, who may have worked harder to understand and explain words to a younger peer than they would have with a same-age peer.

While there may have been effects on target word learning, there were no effects on distal measures of general vocabulary for kindergarteners or comprehension for fourth-grade students. We did not measure general comprehension for kindergarteners or general vocabulary for fourth-grade students due to time constraints. The program may not have had an effect on general vocabulary knowledge since it did not include explicit instruction on vocabulary (National Reading Technical Assistance Center, 2010) or generalizable word learning strategies (Graves, 2006). Future researchers should investigate whether adding these components would make the effects of the intervention on vocabulary more robust. The fact that there were no effects on comprehension may be due to the fact that the intervention did not include explicit instruction on comprehension strategies (Kamil et al., 2008; Shanahan et al., 2010). Future researchers should explore whether adding explicit instruction on comprehension strategies to reading buddies programs would yield effects on comprehension for older and younger students.

Given that the program showed some positive, albeit small, effects, it is worth considering what the potential strengths of the program may be and what might make the program stronger in future iterations. As teachers who implemented the program indicated, strengths of the program include increasing student interaction through the cross-age peer learning model (Topping, 2005), prioritizing vocabulary learning in text and in conversation (National Reading Technical Assistance Center, 2010), and engaging learners through different text types such as video and regular text (Silverman & Hines, 2009). All of these elements may be supportive of vocabulary learning. For teachers, the program required little training and preparation, and took up relatively little instructional time (roughly 10 hr across the school year).

Analyzing teacher responses to the survey about strengths and challenges of the program, teachers felt the program was easy to implement and aligned well with their classroom and school curriculum and testing standards. Teachers indicated that the cross-age format of the program was a strength, and many of them mentioned that the program challenged the fourth-grade students to take on a leadership role. In addition, the multiple modalities the program used to present and reinforce new vocabulary was another strength teachers mentioned repeatedly in their survey responses. One teacher specifically mentioned that the program encouraged students to think actively about words and texts, and provided an authentic context for vocabulary learning.

In addition to targeting language and literacy skills, teachers also reported that the program improved student motivation and engagement in the classroom, which is a finding supported by other reviews of cross-age programs (e.g., Cohen et al., 1982; Gorrell & Keel, 1986). Specifically, multiple teachers mentioned that their students looked forward to the reading buddy sessions and established close relationships with their buddies. Teachers suggested that the varied activities in the lessons helped maintain student engagement over the course of the buddy sessions. Future research on how reading buddies programs support engagement and motivation may show positive results of such programs on academics and beyond (Van Keer & Vanderlinde, 2010).

Limitations

There were several limitations to the present study. Limitations were inherent to the program itself and to the research study we conducted. In an effort not to overburden teachers, the program was kept intentionally short required little teacher training or preparation. While this likely helped with teacher buy-in and fidelity to program implementation, the program may not have been robust enough to have effects on generalizable measures of vocabulary and comprehension. Future researchers should investigate cross-age peer learning programs over a longer period of time and with more teacher involvement and support. Teachers who conducted the program were asked to refrain from deliberately reinforcing concepts from the program at other times of the day so that effects of the program could be attributable to the program itself rather than other activities teachers implemented. Future researchers should explore whether reinforcing concepts of the program outside of the reading buddies sessions would make the program more effective. To limit the testing time required of students, we did not include measures of kindergarten comprehension or Grade 4 generalizable vocabulary knowledge, and we did not include measures of student engagement or motivation. Future researchers should include such measures. Although we included a control group of students that were compared to the intervention students on measures of vocabulary and comprehension, observations of the control classrooms were not

conducted. Therefore, it is not known what vocabulary and comprehension instruction students in the control classroom received. The program included multiple components and activities. Given the present study design and the type of data collected in the study, it is impossible to determine which aspects of the program led to student target word learning. Future researchers should compare different versions of the program that include or exclude certain components to determine which components are most important for program effectiveness. Additionally, future researchers should include qualitative data collection during reading buddy sessions to help investigators understand the underlying mechanisms that lead to student vocabulary learning and determine how the program could be improved to lead to more generalizable effects. Finally, future researchers should further explore the role of teacher perceptions and implementation on program effectiveness and the effects of the program for different groups of students (e.g., ELL, low socioeconomic status, or special education students).

Conclusion

The present study suggests that the MSRB program holds promise for kindergarten and fourth-grade students in the development of vocabulary knowledge. In classrooms with predominantly teacher-led vocabulary and comprehension instruction, students have relatively little time to use words they are learning in conversation with others. Encouraging students to talk about words and texts through cross-age peer learning programs such as the MSRB program may help children internalize the words they are learning so that they can use them to understand and talk about content and text in school. Much more research is needed on cross-age peer learning programs, but, given the current emphasis on vocabulary and comprehension in elementary schools, it is worth further exploration of reading buddies programs to determine how they may contribute to promoting students' word knowledge and ability to understand content and texts in school.

References

Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness*, *5*, 160–188.

Babicki, L., & Luke, S. (2007). Oh, and we do reading too. *Montessori Life*, *19*, 46.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., … Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, *107*, 251–271.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford.

Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Kromrey, J. D., & Ferron, J. M. (2014). How low can you go? An investigation of the influence of sample size and model complexity on point and interval estimates in two-level linear models. *Methodology*, *10*, 1–11.

Biemiller, A. (2004). Teaching vocabulary in the primary grades: Vocabulary instruction needed. In J. F. Bauman & E. J. Kame'enui (Eds.), *Vocabulary instruction: Research to practice*. New York, NY: Guildford Press.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, *49*, 278–293.

Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P. C., Tucker, B. J., Cheung, A., & Gifford, R. (2008). Technology infusion in success for all: Reading outcomes for first graders. *Elementary School Journal*, *109*, 1–15.

Christ, T., & Wang, X. C. (2012). Young Children's opportunities to use and learn theme-related vocabulary through buddy "Reading". *Literacy Research and Instruction*, *51*, 273–291.

Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*, 237–248.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*, 934–945.

Dalton, B., Proctor, C. P., Uccelli, P., Mo, E., & Snow, C. E. (2011). Designing for diversity the role of reading strategies and interactive vocabulary in a digital reading environment for fifth-grade monolingual english and bilingual students. *Journal of Literacy Research*, *43*, 68–100.

Davenport, S., Arnold, M., Lassmann, M., & Lasmann, M. (2004). The impact of cross-age tutoring on reading attitudes and reading achievement. *Reading Improvement*, *41*, 3–12.

DePaulo, B. M., Tang, J., Webb, W., Hoover, C., Marsh, K., & Litowitz, C. (1989). Age difference in reactions to help in a peer tutoring context. *Child Development*, *60*, 423–439.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Minneapolis, MN: NCS Pearson.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248.

Fuchs, D., & Fuchs, L. S. (2005). Peer-assisted learning strategies promoting word recognition, fluency, and reading comprehension in young children. *The Journal of Special Education*, *39*, 34–44.

Fuchs, D., Fuchs, L. S., & Burish, P. (2000). Peer-assisted learning strategies: An evidence-based practice to promote reading achievement. *Learning Disabilities Research & Practice*, *15*, 85–91.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*, 174–206.

Gartner, A., Kohler, M., & Riessman, F. (1971). *Children teach children: Learning by teaching*. New York, NY: Harper & Row.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasiexperimental research in special education. *Exceptional Children*, *71*, 149–164.

Glaser, B., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.

Good, R. H. III, Kaminski, R. A., Cummings, K., Dufour-Martel, C., Petersen, K., Powell-Smith, K., … Wallin, J. (2010). *DIBELS next assessment manual*. From https://dibels.org/next/ (accessed April 8, 2015).

Goodson, B., Wolf, A., Bell, S., Turner, H., & Finney, P. B. (2011). *Effectiveness of a program to accelerate vocabulary development in kindergarten (VOCAB): First grade follow-up impact report and exploratory analyses of kindergarten impacts* (NCEE 2012-4009). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gorrell, J., & Keel, L. (1986). A field study of helping relationships in a cross-age tutoring program. *Elementary School Guidance & Counseling*, *20*(4), 268–276.

Graves, M. F. (2006). *The vocabulary book: Learning and instruction*. New York, NY: Teachers College Press.

International Reading Association. (2012). *Literacy implementation guidance for the ELA Common Core State Standards*. Retrieved from http://www.reading.org/libraries/association-documents/ira_ccss_guidelines.pdf.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* (NCEE #2008–4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Kamil, M. L., Intrator, S., & Kim, H. (2000). The effects of other technologies on literacy and literacy learning. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 771–788). Mahwah, NJ: Erlbaum.

Kelly, M., Moore, D. W., & Tuck, B. F. (1994). Reciprocal teaching in a regular primary school classroom. *The Journal of Educational Research*, *88*, 53–61.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*, 983–997.

Kenward, M. G., & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis*, *53*, 2583–2595.

Kim, Y. S., Wagner, R. K., & Foster, E. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific Studies of Reading*, *15*, 338–362.

Kim, Y. S., Wagner, R. K., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from Grade 1 to Grade 2. *Journal of Experimental Child Psychology*, *113*, 93–111.

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology*, *90*, 134–152.

Klingner, J. K., & Vaughn, S. (1998). Using Collaborative Strategic Reading. *Teaching Exceptional Children*, *30*(6), 32.

Klingner, J. K., & Vaughn, S. (1999). Promoting reading comprehension, content learning, and English acquisition though Collaborative Strategic Reading (CSR). *Reading Teacher*, *52*, 738.

Klingner, J. K., & Vaughn, S. (2000). The helping behaviors of fifth graders while using collaborative strategic reading during ESL content classes. *TESOL Quarterly*, *34*, 69–98.

Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *Elementary School Journal*, *96*, 275–293.

Kutnick, P., Blatchford, P., & Baines, E. (2002). Pupil groupings in primary school classrooms: Sites for learning and social pedagogy? *British Educational Research Journal*, *28*, 187–206.

Leko, M. M. (2014). The value of qualitative methods in social validity research. *Remedial And Special Education*, *35*, 275–286.

Leung, C., Silverman, R., Nadakumar, N., & Hines, S. (2011). Difficulty of words encountered in first-grade basal readers: A Rasch model for preschool ELL and monolingual English speakers. *American Educational Research Journal*, *48*(2), 421–461.

Lindo, E. J., & Elleman, A. M. (2010). Social validity's presence in field-based reading intervention research. *Remedial and Special Education*, *31*, 489–499.

Lowery, R. M., Sabis-Burns, D., & Anderson-Brown, S. (2008). Book buddies: Kindergarteners and fifth graders explore books together. *Dimensions of Early Childhood*, *31*(3), 31–38.

Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. *The Elementary School Journal*, *90*, 469–484.

Mahdavi, J. N., & Tensfeldt, L. (2013). Untangling reading comprehension strategy instruction: Assisting struggling readers in the primary grades. *Preventing School Failure: Alternative Education for Children and Youth*, *57*, 77–92.

Manyak, P. C., Von Gunten, H., Autenrieth, D., Gillis, C., Mastre-O'Farrell, J., Irvine-McDermott, E., … Blachowicz, C. L. (2014). Four practical principles for enhancing vocabulary instruction. *Reading Teacher*, *68*, 13–23.

McMaster, K. L., Fuchs, D., & Fuchs, L. S. (2007). Promises and limitations of peer-assisted learning strategies in reading. *Learning Disabilities: A Contemporary Journal*, *5*, 97–112.

McNeish, D. M., & Stapleton, L. M. (2014). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*. doi:10.1007/s10648-014-9287-x.

Meddaugh, S. (1995). *Martha speaks*. Boston, MA: Houghton Mifflin Harcourt.

Morel, J. G., Bokossa, M. C., & Neerchal, N. K. (2003). Small sample correction for the variance of GEE estimators. *Biometrical Journal*, *45*, 395–409.

National Center for Education Statistics. (2013). *The Nation's report card: A first look: 2013 mathematics and reading* (NCES 2014-451). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

National Reading Technical Assistance Center. (2010). *A review of the current research on vocabulary instruction*. Portsmouth, NH: RMC Research Corporation.

Neuman, S. B. (1992). Is learning from media distinctive? Examining children's inferencing strategies. *American Educational Research Journal*, *29*(1), 119–140.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition & Instruction*, *1*, 117.

Proctor, C. P., Dalton, B., & Grisham, D. L. (2007). Scaffolding english language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research*, *39*, 71–93.

Proctor, P., Silverman, R., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing: An Interdisciplinary Journal*, *25*(7), 1635–1664.

Puzio, K., & Colby, G. T. (2013). Cooperative learning and literacy: A meta-analytic review. *Journal of Research on Educational Effectiveness*, *6*, 339–360.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*, 240–257.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, *64*, 479–530. doi:10.2307/1170585

Saenz, L. M., Fuchs, L. S., & Fuchs, D. (2005). Peer-assisted learning strategies for english language learners with learning disabilities. *Exceptional Children*, *71*, 231.

Saltzberg, B. (2010). *Star of the week*. Somerville, MA: Candlewick Press.

Schneider, R. B., & Barone, D. (1997). Cross-age tutoring. *Childhood Education*, *73*, 136–143.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Silverman, R., & Hines, S. (2009). The effects of multimedia enhanced instruction on the vocabulary of English language learners and non-English language learners in pre-kindergarten through second grade. *Journal of Educational Psychology*, *101*(2), 305–314.

Silverman, R. D., Proctor, C. P., Harring, J. R., Doyle, B., Mitchell, M. A., & Meyer, A. G. (2014) Teachers' instruction and students' vocabulary and comprehension: An exploratory study with English monolingual and Spanish-English bilingual students in grades 3-5. *Reading Research Quarterly*, *49*, 31–60.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, *21*, 43–69.

Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, *50*, 1171–1177.

Topping, K. (2005). Trends in peer learning. *Educational Psychology*, *25*, 631–645.

Topping, K., Campbell, J., Douglas, W., & Smith, A. (2003). Cross-age peer tutoring in mathematics with seven-and 11-year-olds: Influence on mathematical vocabulary, strategic dialogue and self-concept. *Educational Research*, *45*, 287–308.

Topping, K. J., Peter, C., Stephen, P., & Whale, M. (2004). Cross-age peer tutoring of science in the primary school: Influence on scientific language and thinking. *Educational Psychology*, *24*, 57–75.

Topping, K. J., Thurston, A., McGavock, K., & Conlin, N. (2012). Outcomes and process in reading tutoring. *Educational Research*, *54*, 239–258.

Van Keer, H., & Vanderlinde, R. (2010). The impact of cross-age peer tutoring on third and sixth graders' reading strategy awareness, reading strategy use, and reading comprehension. *Middle Grades Research Journal*, *5*, 33–45.

Van Keer, H., & Verhaeghe, J. P. (2005). Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions. *The Journal of Experimental Education*, *73*, 291–329.

Vaughn, S., Klingner, J. K., & Bryant, D. P. (2001). Collaborative strategic reading as a means to enhance peer-mediated instruction for reading comprehension and content-area learning. *Remedial and Special Education*, *22*, 66–74. doi:10.1177/074193250102200201

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wagner, R. K., Torgesen, J., Rashotte, C. A., & Pearson, N. (2010). *Test of sentence reading efficiency and comprehension*. Austin, TX: Pro-Ed.

WGBH Boston (Producer). (2008). *Martha speaks* [Television series]. Boston, MA: WGBH Boston.

WGBH Educational Foundation. (2008). *Martha Speaks Reading Buddies Program*. Boston, MA: WGBH Boston.

Zhang, J., Anderson, R., & Nguyen-Jahiel, K. (2013). Language-rich discussions for English language learners. *International Journal of Educational Research*, *58*, 44–60.

Appendix

Teacher survey questions, codes, and sample quotes

Survey question	Codes	Sample teacher quotes
(1) Please rate how difficult the Martha Speaks Program was to implement: easy, fairly easy, moderate, hard, very hard.	Easy, fairly easy, moderate, hard, very hard	The directions were very clear and all the materials were provided. I was very impressed with the level of organization and routine that the program had for the students. They enjoyed it very much.
(2) Were there any barriers to program implementation? Please check yes or no and please explain.	Time in schedule, pacing within lessons, use of technology, no barrier	Scheduling was tricky because our 4th grade schedule is very different from our kindergartners- this made for a lot of work. The only barrier we faced to implement was the technology. However, once the technician of the school got involved we were able to bypass this issue.
(3) Did the program fit easily into your classroom lesson plans? Please check yes or no and please explain.	Alignment to curriculum, no alignment	The program worked well with the classroom lesson plans. The students were getting vocabulary, comprehension, and writing practice. This program also allowed the students to work on collaboration skills. I really loved how the book matched with the video. The students were able to make a quick connection and it relates with our Common Core standards.
(4) Do you feel that your students benefited from participating in the program? Please check yes or no and please explain.	Vocabulary words, work cooperatively, had fun, no benefit	Students gain vocabulary knowledge, listening comprehension, cooperative learning skills, and it made learning fun!They practiced responsibility and maturity. Also, by reading and teaching the lessons, they too gained academic content.
(5) What were the strengths of the program?	Cross-age learning, use of multiple-modalities, enriched vocabulary, alignment with standards	I felt that this program was friendly for all the different types of learners. The students enjoyed that each lesson had a variety of skills and activities. The program let them view video, read, practice vocabulary, play games, talk, communicate, and write their ideas. They also enjoyed the making using their creativity. The strengths of the program would be how it is such an easy process to follow. The books are fantastic choices that match the video and the activities were appropriate for each session. My students looked forward to the activity each day. I like the structure of each lesson. The clips really did a good job of teaching the vocabulary words. The materials were spot on!
(6) Did the program help with any of the following? Check all that apply: vocabulary, comprehension, fluency, motivation, engagement	Vocabulary, comprehension, fluency, motivation, engagement	Teacher comments are not applicable for this question.
(7) What suggestions do you have for improving the program?	Begin earlier, text length/level, more substance to game and journal writing, more hands-on activities, reduce Grade 4 practice time	Fit better in the timeframe. There needs to be more meat to the games and journal writing. Would like more hands on activities for the little ones.