Early Childhood Research Quarterly 31 (2015) 47–61

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Effects of automated Tier 2 storybook intervention on vocabulary and comprehension learning in preschool children with limited oral language skills

Elizabeth Spencer Kelley[a,*], Howard Goldstein[b], Trina D. Spencer[c], Amber Sherman[d]

[a] Department of Communication Science and Disorders, University of Missouri – Columbia, 423 Lewis Hall, Columbia, MO 65211, United States
[b] College of Behavioral & Community Sciences, University of South Florida, 13301 Bruce B. Downs Blvd., MHC 1121, Tampa, FL 33612-3807, United States
[c] Institute for Human Development, Northern Arizona University, PO Box 5630, Flagstaff, AZ 86011, United States
[d] Crane Center for Early Childhood, The Ohio State University, 1945 N High Street, Room 353, Columbus, OH 43210, United States

ARTICLE INFO

Article history:
Received 23 January 2014
Received in revised form 3 December 2014
Accepted 29 December 2014
Available online 8 January 2015

Keywords:
Vocabulary
Preschool
Comprehension
Language intervention
Response to intervention

ABSTRACT

This early efficacy study examined the effects of an automated storybook intervention designed to promote school readiness among at-risk prekindergarten children. Story Friends is a small-group intervention in which vocabulary and question-answering lessons are embedded in a series of storybooks. A randomized group design with an embedded single-case experimental design was used to examine treatment effects. Eighteen children from public prekindergarten programs serving families with low income were randomly assigned to the Story Friends treatment or a business-as-usual comparison. Participants in both groups completed measures of vocabulary and comprehension approximately monthly. Participants in the treatment group completed measures of instructional content for each book as part of the embedded single-case experimental design. Story Friends participants had significantly higher scores on measures of vocabulary than the comparison group and effect sizes were large, whereas more modest effects were shown for comprehension measures. Observations of treatment fidelity indicate that this intervention has the potential to be implemented with high fidelity in preschool classrooms. Results show a feasible means of teaching pre-K children challenging vocabulary that has the potential to facilitate later literacy development.

Introduction

Children enter early childhood programs with diverse early language and literacy experiences. A substantial number of children, including those from families with low socioeconomic status, have limited oral language skills that place them at risk for later reading disabilities. For example, Qi, Kaiser, Milan, and Hancock (2006) reported that a group of preschoolers enrolled in Head Start had a group mean of approximately 1.5 standard deviations below the normative mean on a standardized measure of receptive vocabulary, the Peabody Picture Vocabulary Test – Third Edition (PPVT-III; Dunn & Dunn, 1996). Oral language skills, including vocabulary, in early childhood predict future reading ability (Lonigan, Schatschneider, & Westburg, 2008), placing preschool children with limited oral language skills at high risk of reading disabilities in later school years.

Response to intervention

Multi-tiered instruction, a key component of response to intervention (RtI) models, is a promising approach for preventing reading disabilities. RtI models have been implemented widely in the elementary years (Berkeley, Bender, Gregg Peaster, & Saunders, 2009) and are an emerging practice in early childhood settings (Greenwood et al., 2011; VanDerHeyden, Snyder, Broussard, & Ramsdell, 2008). Children who receive prompt instruction to remediate academic deficits within a multi-tier framework may experience improved academic outcomes. Specifically, effective tiered oral language and literacy instruction in early childhood may improve skills of young children and, thus, prevent future reading disabilities.

In an RtI model, different tiers of instruction, often three, are provided to children based on individual needs (Gersten et al., 2008).

* Corresponding author. Tel.: +1 573 882 2981.
E-mail addresses: kelleyes@health.missouri.edu (E.S. Kelley), hgoldstein@usf.edu (H. Goldstein), trina.Spencer@nau.edu (T.D. Spencer), Sherman.156@osu.edu (A. Sherman).

http://dx.doi.org/10.1016/j.ecresq.2014.12.004
0885-2006/

In Tier 1, a high-quality general education curriculum is provided to all children. Information from screening or progress monitoring measures is used to identify children who are not making adequate progress in Tier 1 and who may benefit from supplemental instruction in a Tier-2 arrangement. Children who fall well behind peers, and for whom Tier 2 is not sufficient, may receive intensive, individualized instruction in Tier 3. Tiers 2 and 3 can be differentiated from Tier 1 by the frequency and duration of instruction, the instructional approach (e.g., systematic, explicit), and the delivery arrangement (e.g., small group or individual; Fuchs & Fuchs, 2006). Most often, high quality Tier-2 programming is characterized by systematic, supplemental, targeted instruction administered in small groups.

Effective implementation of RtI models in early childhood settings necessitates empirically supported options at all three tiers. However, researchers have reported a concern about the general effectiveness and quality of Tier-1 instruction in early childhood settings (Greenwood et al., 2012; Justice, Mashburn, Hamre, & Pianta, 2008; Justice, Mashburn, Pence, & Wiggins, 2008), which creates an important challenge to the effectiveness of RtI models. Hence, investigators have sought to improve Tier-1 instruction in early childhood settings (Diamond & Powell, 2011; Dickinson & Caswell, 2007; Justice, Mashburn, Hamre, et al., 2008; Justice, Mashburn, Pence, et al., 2008). However, there also is a need for high-quality Tier-2 and Tier-3 interventions for young children with learning needs.

For children with limited oral language skills, supplemental intervention may help prevent academic problems. Prior vocabulary knowledge is a predictor of success in vocabulary intervention studies, as children who begin intervention with low vocabulary are less likely that peers with higher vocabulary to learn words (Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009, Coyne, Simmons, Kame'enui, & Stoolmiller, 2004; Penno, Wilkinson, & Moore, 2002). Few studies have examined tiered approaches for improving the oral language of young children with limited oral language skills (Loftus, Coyne, McCoach, & Zipoli, 2010; Pullen, Tuckwiller, Konold, Maynard, & Coyne, 2010; Zucker, Solari, Landry, & Swank, 2013). Loftus et al. (2010) delivered a tiered vocabulary intervention to kindergartners identified as at-risk because of low vocabulary scores (standard scores between 40 and 91 on the PPVT-III; Dunn & Dunn, 1996). Participants learned more about those words if they received both a first tier of classroom-based instruction and a second tier of supplemental instruction than if received only the first tier of instruction. Using a similar approach with a between-subjects design, Pullen et al. (2010) reported that at-risk children who received a second tier of vocabulary instruction made gains in vocabulary, whereas at-risk peers who received only the first tier did not.

Characteristics of effective oral language interventions

Several recent meta-analyses and research syntheses have reported moderate-to-large effects of various interventions on oral language skills of young children. Those interventions included shared book reading, language enhancement, and vocabulary interventions (Elleman, Lindo, Morphy, & Compton, 2009; Fischel & Landry, 2008; Lonigan, Shanahan, & Cunningham, 2009; Marulis & Neuman, 2010; Mol, Bus, & de Jong, 2009; Mol, Bus, de Jong, & Smeets, 2008). Within this evidence of positive intervention effects, there is substantial variability in the magnitude of effects. For example, in the 67 vocabulary intervention studies included in Marulis and Neuman (2010) review, effect sizes ranged from −0.10 to +2.13. To substantially improve outcomes for children with limited oral language skills as part of tiered models of instruction, it is important that interventions produce strong effects.

Based on RtI research with school-age students, several features of effective supplemental interventions have been suggested. Gersten et al. (2008) recommend that Tier-2 intervention should be implemented in small groups, target critical reading-related skills, and include explicit instruction with multiple opportunities for student practice. Foorman and Torgesen (2001) asserted that effective instruction for children at-risk of reading failure should be explicit, intensive, and supportive (e.g., include scaffolding to help children acquire new skills).

When studies of oral language intervention are examined, the same characteristics of effective interventions emerge. Marulis and Neuman (2010) concluded that only vocabulary interventions using explicit teaching strategies produce large effects. Children learn more words and more about those words when provided explicit instruction compared to when children are simply exposed to words (Brett, Rothlein, & Hurley, 1996; Coyne, McCoach, & Kapp, 2007; Elley, 1989; Justice, Meier, & Walpole, 2005; Penno et al., 2002; Robbins & Ehri, 1994; Sénéchal, 1997).

Effective instruction is intensive and interactive. Interventions in which children receive repeated exposure to vocabulary instruction is more effective than limited exposure (Beck & McKeown, 2007; Coyne et al., 2007). Oral language skill instruction has been shown to be more effective when it is interactive (Mol et al., 2008; Whitehurst et al., 1994). Components of interactive instruction include opportunities for active responding by children (Greene-Brabham & Lynch-Brown, 2002; Sénéchal, Thomas, & Monker, 1995), modeled feedback (van Kleeck, van der Woude, & Hammett, 2006) and including open-ended questions (Wasik & Bond, 2001).

These characteristics are evident in the model of robust vocabulary instruction advanced by Beck & McKeown (2007) and Beck, McKeown, and Kucan (2013). Robust vocabulary instruction includes recommendations for the types of words most appropriate for explicit, intensive instruction as well as recommendations for the instructional approach. Recommended target vocabulary words are 'Tier-2' words that are sophisticated, high-utility words (*required*, *maintain*; Beck et al., 2013). Instruction is designed to develop deep understanding of these words. Words are presented with explicit instruction that includes child-friendly definitions and multiple examples and contexts to provide information about word meanings. The positive effects of this type of extended, explicit instruction on the vocabulary knowledge of young children have been well documented (Coyne et al., 2009; Justice et al., 2005; Penno et al., 2002; Pollard-Durodola et al., 2011). Participants in these studies demonstrated learning of target vocabulary but rarely demonstrated improvements on generalized measures of oral language (Coyne et al., 2010).

Several research groups have examined inferential language as a contributor to comprehension. Inferencing ability predicts later listening comprehension (Kendeou, Bohn-Gettler, White, & Van Den Broek, 2008; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012) and studies of inferential question use by teachers and parents indicate that these types of questions may result in children's use and understanding of more sophisticated, abstract language (Tompkins, Zucker, Justice, & Binici, 2013; van Kleeck, Gillam, Hamilton, & Cassandra, 1997; Zucker, Justice, Piasta, & Kaderavek, 2010). Few studies have examined interventions to teach such comprehension skills to young children. van Kleeck et al. (2006) examined the effects of a scripted storybook intervention that targeted inferential as well as literal questions. Inferential questions are related to information that is not directly stated in the text or illustrations of the story (e.g., predictions, questions about character emotions). Preschool children with language impairments demonstrated larger gains in literal and inferential language skills relative to a comparison group. Desmarais, Nadeau, Trudeau, Filiatrault-Veilleux, & Maxès-Fournier (2013) reported positive effects of a similar intervention, although the lack of a control group compromises the interpretation of their findings.

Story Friends intervention

The characteristics discussed were incorporated into the development of *the Story Friends* curriculum, the focus of the current study. The curriculum was designed as a Tier-2 vocabulary and comprehension intervention that included key components of explicit teaching, repeated exposure, and interactive instruction that incorporated multiple opportunities to respond. In the context of an RtI model, it is important to consider the demands placed on educational staff to implement multiple tiers of intervention in several domains with high treatment fidelity (Ukrainetz, 2006). Thus, the design of Story Friends was based on the premise that pre-K teachers need intervention options that do not require extensive time for training, preparation, or implementation. Hence, we developed an automated intervention. Automated methods of delivery in which instruction is provided by prerecorded or computer-assisted means can be delivered with high fidelity while minimizing demands on educational staff. Automated instruction also can ensure consistent delivery of instruction and repeated experiences with lessons can provide intensity as children receive many more opportunities to respond in comparison to the turn-taking that is more common in teacher-led small group instruction. Mayer (2005) proposes other potential advantages stemming from a cognitive theory of multimedia learning. He argues that more learning takes place when information is presented in multiple formats (e.g., words and pictures) rather than with words alone. Using prerecorded narration and accompanying interactive storybook activities, multimedia components of instruction may serve to highlight key events in a story or provide information about a word's meaning.

Reviews of the effects of computer-based instruction have indicated an overall positive, but small, effect on learning of early language and literacy skills (Blok, Oostdam, Otter, & Overmaat, 2002; van Daal & Sandvik, 2013). Computer-based programs have improved phonological awareness skills (Lonigan, Phillips, Cantor, Anthony, & Goldstein, 2003; Segers & Verhoeven, 2004) and vocabulary (Segers & Verhoeven, 2003). Interactive electronic books have been used to teach early language and literacy skills, including vocabulary, to young children (Shamir, Korat, & Fellah, 2010; Verhallen & Bus, 2010; Verhallen, Bus, & de Jong, 2006). Verhallen et al. (2006) reported that children learned more vocabulary from multimedia, interactive storybooks than from static versions.

Iterative development of Story Friends

The development of *Story Friends* has followed an iterative process. Spencer et al. (2013) investigated an earlier iteration of Story Friends using a repeated acquisition single-case design across intervention targets to examine the extent to which Story Friends improved vocabulary and comprehension skills of nine pre-K students. Consistent improvements in vocabulary and modest gains in comprehension indicated that Story Friends had potential for effective use in early childhood classrooms.

Findings from studies of earlier versions of *Story Friends* (Greenwood et al., in press; Spencer et al., 2013, unpublished pilot studies) informed revisions to program materials, intervention procedures, and assessments. Revisions were made to target vocabulary, story questions, embedded lessons, assessments procedures, and implementation materials. For example, five words that few children learned (e.g., *unusual*) were replaced and embedded lessons were revised for seven of the words. For example, we found that lessons that included photograph examples, as opposed to illustrations, were more successful in providing clear contexts for word meanings. To increase opportunities for practice of previously taught words, review books were created that repeated vocabulary lessons after every three instructional books.

We also replaced 18 of 27 story questions and accompanying embedded lessons. For example, because children made the fewest gains on pre-story questions (e.g., *What do you think will happen in this story?*) these questions were replaced with questions embedded in the story. Because we found that participants could frequently answer literal questions at pretest; these questions were replaced with inferential questions.

Although we had used weekly mastery monitoring probes to gather detailed information about participant learning on instructional targets, we anticipated that such frequent testing would be impractical when teachers implemented the intervention. To address this concern, we created two new measures that would be administered approximately monthly: Unit Vocabulary Tests and the Assessment of Story Comprehension. Finally, to address the goal of high fidelity implementation, we revised intervention manuals, training materials, and scoring guidelines.

Thus, the current study extended the Spencer et al. (2013) study by implementing a revised and refined intervention with a larger group of children. The study design was a randomized group-design study with an embedded repeated acquisition design. The latter single-subject experimental design provided careful examination of the effects of the intervention (e.g., learning of individual targets by each participant). The group design served as an early efficacy study in preparation for a clinical trial by allowing the research team to refine procedures for implementation and measurement and to estimate effect sizes.

The following research questions were addressed:

1. To what extent does automated vocabulary intervention embedded in prerecorded storybooks improve vocabulary knowledge of prekindergarten children with limited oral language skills?
2. To what extent does automated question-answering intervention embedded in prerecorded storybooks improve the ability of preschool children with limited oral language skills to answer questions about stories?

We hypothesized that preschool participants would demonstrate increases in both vocabulary knowledge and question-answering skills. Based on our previous studies and existing research, large differences between treatment and comparison groups at posttest were predicted.

Overview of the intervention

Story Friends intervention involves brief instructional lessons embedded in prerecorded storybooks. The instructional lessons address vocabulary words and answering questions about the story. While listening to the story, children interact with an attractive storybook and are encouraged to respond to the narrator. The Story Friends materials included storybooks, prerecorded audio, mp3 players, and headphones. The research team wrote stories featuring a group of animal characters, the Jungle Friends. Stories were written with a focus on common childhood experiences such as a first day of school or a trip to the dentist's office. Each storybook was between 9 and 11 pages long and story text rhymed. A professional artist created the illustrations and a professional narrator recorded the story text and embedded lessons.

Story Friends includes a total of 13 storybooks: one introductory book, nine instructional books, and three review books. The instructional and review books are organized into three units consisting of three instructional books and a review book. The introductory book, *Meet the Jungle Friends* familiarized children with the characters in the stories and with the interactive activities (e.g., turning pages, lifting flaps, and responding to questions).

For the purposes of the study, two versions of each book were created. One version included the story text and illustrations only

Table 1
Sample instructional language for embedded lessons.

Embedded vocabulary lessons	
Instructional language component	Example for *Thrilled*
Connection to the story context	*Wow! The Jungle Friends are thrilled! They are excited to go to the carnival*
Opportunity to repeat the word	*Thrilled. Say thrilled*
Simple definition	*Thrilled means excited*
Child-friendly context	*When are you thrilled? What about when you get a present. Or your friends come over to play? I bet that makes you feel excited*
Activity	*Now lift the flap.* [Picture of two young boys with party hats and a birthday cake.] *They are at a birthday party! They are thrilled. They are excited*
Opportunity to say the definition	*Tell me, what does thrilled mean?* [pause] *Excited!*
Basic concept word lesson	
Instructional language component	Example for *high*
Connection to the context of the story	*Oh my goodness! The ball was stuck high in the tree!*
Opportunity to say the word(s)	*High. Say high*
Activity	*Put your hand way up in the air*
Story question lesson	
Instructional language component	Example for *The other kids laughed at Ellie Elephant. Why did they laugh at Ellie?*
Model of appropriate response	*Because she was new!*
"Think Aloud" explanation of response	*Ellie broke her chair, and no one could lift her on the see-saw. The other kids didn't know her yet, so they laughed at her*

(e.g., the 'plain' version) and the audio narration included just the story text. The 'plain' versions provided exposure to the story without providing embedded lessons, so the children could hear the story before being asked to respond to questions about the story. The second version of the books included story text, illustrations, and pages with pictures and activities for embedded lessons (e.g., the 'intervention' version). The narration included the story text and the embedded instructional lessons for vocabulary words and story questions. Plain storybooks were between 4 and 7 min long and instructional storybooks were between 9 and 11 min long.

Three review books were designed to provide a brief review of embedded vocabulary words. The review books included one lesson for each target vocabulary word in the unit for a total of six lessons. The review books did not include a story; instead review books were repetitions of the lessons in the instructional books. Pictures, instructional language, and accompanying audio were identical to the intervention lessons. Review books were 4–6 min.

In each instructional book, embedded lessons were included for two challenging vocabulary words, one or two basic concept words, and three story questions. Thus, across the series of books, a total of 18 challenging words, 15 basic concept words, and 27 story questions were taught. Embedded lessons were brief interruptions to the story and included instruction, opportunities for children to respond, and models of appropriate responding. Lessons followed a consistent format with systematic instructional language. All embedded lessons were delivered via prerecorded audio. See Table 1 for sample embedded lessons.

Challenging vocabulary words

Challenging vocabulary words were selected to represent 'Tier-2′ words in the model of robust vocabulary instruction recommended by Beck et al. (2013). Thus, we selected sophisticated, high-utility words and looked for words that represented concepts that preschool children would know and that would allow for greater precision in expression (e.g., *soaked* as a more precise word for *wet*). We targeted challenging words that were unlikely to be known by children with limited vocabulary but were likely known by peers with large vocabularies or older children. We operationalized our criteria for selecting words (Spencer, Goldstein, & Kaminski, 2012) to include: (a) the word was unlikely to be in the lexicon of pre-K children with limited vocabulary, (b) the word was likely to occur relatively frequently in the conversation of adult speakers, (c) the word could be defined simply with a child-friendly definition (e.g., the definition of *grin* was *to smile*; the definition of *ridiculous* was *silly*), and d) the word could be supported in the context of the storybook.

Each word was targeted in two lessons: one lesson during the story and one lesson immediately following the story. The first lesson included: (a) a connection to the context of the story, (b) an opportunity to repeat the word, (c) an explicit simple definition, (d) an opportunity to say the word in response to the definition, (e) an example of the word's use in a child-friendly context, and (f) an opportunity to say the definition. These lessons were 1–1½ min in length and included a minimum of four opportunities for children to respond. The second lesson was slightly shorter and included: (a) an opportunity to repeat the word, (b) the explicit simple definition, (c) a second example of the word's use in a child-friendly context, and (d) an opportunity to say the definition. These lessons were 15–30 s in length and included a minimum of three opportunities to respond. Embedded lessons included models of appropriate responses and verbal encouragement. For example, the question, "*what does enormous mean?*" would be followed by a short pause (~3 s) and then the narrator would respond with the appropriate response "*really big*" and provide encouragement "*great job!*" Because the lessons were all prerecorded, the feedback provided by the audio was the same regardless of participant behavior; children heard the correct answer and encouragement whether or not they responded.

Basic concept words

For each book, one or two basic concept words (e.g., *high, low; many, few)* were selected from early childhood assessments such as the Boehm Test of Basic Concepts (Boehm, 2000). Targeted words were age appropriate (e.g., from the 4-year-old list of the Boehm) and could be easily embedded in the story text. Each basic concept word was addressed in a lesson embedded in the story and again immediately after the story. Lessons for the basic concept words were brief and included an example of the concept word using a picture or illustration (e.g., a ball *high* in the tree) and an opportunity to say the words (e.g., *The ball was stuck high in the tree! High. Say high.*) As in previous studies (Spencer et al., 2013), participants knew many of the basic concept words at pretest (pretest means of 11.4 out of 15, SD = 3.35). Given the high levels of pretest knowledge, we did not examine learning of basic concept words as an outcome of the intervention. Instead, the basic concept word lessons served as an opportunity for participants' to be successful during the storybooks.

Story questions

Three story questions were created for each story. One question was embedded during the story; two questions occurred immediately following the story. Story questions were inferential rather than literal questions (van Kleeck, 2006; van Kleeck et al., 2006) and included questions about character emotions (e.g., *Why was Ellie happy?*), character actions (e.g., *Why did Leo tell Marquez to turn around?)*, post-story predictions (e.g., *Do you think the Jungle Friends will go to the beach again? Why/why not?)*, and questions that made a connection between the story and the child's life (*In this story, the Jungle Friends played tag. What game do you like to play?*). For almost all questions, appropriate responses to the questions would include

information from the story. For example, for the question *Why was Ellie excited?*, the story text included several reasons that Ellie might be happy (e.g., meeting new friends, the first day of school). Embedded lessons for story questions included a model of an appropriate response and a 'think aloud' that explained the response. For example, a modeled answer, "*Ellie is excited because it is her first day of school!"* and a 'think aloud,' "*She gets to ride the school bus and meet her new teacher! That sounds fun!"* Embedded lessons for story questions were less than a minute long and included one opportunity for children to respond.

Method

Participants

Participants were 18 children, 11 girls and 7 boys, with a mean age of 4 years, 6 months recruited from three prekindergarten classrooms primarily serving children from families with low income. All children who had parental permission for participation were screened.

This study was conducted as part of a larger effort to examine RtI models in early childhood settings (Center for Response to Intervention in Early Childhood, 2009). One component of that effort was the development and evaluation of screening measures to identify candidates for Tier-2 and Tier-3 interventions. To address this purpose, we administered two Individual Growth and Development Indicators (IGDIs 2.0): Picture Naming and Which One Doesn't Belong. The IGDIs are currently under development by researchers at the University of Minnesota (Bradfield et al., 2013). For the Picture Naming IGDI, children were presented with 15 cards with photographs of objects (e.g., baby) and asked to label them orally. For the Which One Doesn't Belong IGDI, children were presented 15 cards with photographs of three objects each (e.g., car, truck, baby) and asked to point to the item that did not belong. For both measures, cards were presented one at a time and there was no time limit. Norm-referenced cut points were not yet available for these measures. However, children scoring less than two on the Picture Naming IGDI were unlikely to have sufficient vocabulary to benefit from the intervention, which targeted sophisticated vocabulary. Two children were eliminated based on this cut off; these children may have been better candidates for a Tier-3 intervention that targeted more basic vocabulary (e.g., common nouns and verbs). Group means on the IGDIs were 5.22 for Picture Naming and 4.78 for Which One Doesn't Belong.

The PPVT-IV (Dunn & Dunn, 2007) served as the primary basis for selecting participants. The PPVT-IV is an individually administered, norm-referenced measure of receptive vocabulary. For each item, participants are presented with a plate of four pictures and asked to select the picture that represents the target word. The PPVT-IV has a normative mean of 100 and a standard deviation of 15. The PPVT-IV has satisfactory evidence of validity and reliability (e.g., strong correlations with other measures of vocabulary, test-retest reliability correlation of 0.93; Dunn & Dunn, 2007). Children were eligible to participate if scores on the PPVT-IV indicated mild-moderate delays in receptive vocabulary (i.e., standard score less than 90). The group mean on the PPVT-IV was 83.44 (SD = 4.02).

To further describe the oral language skill of participants, we administered the core language subtests of the Clinical Evaluation of Language Fundamentals Preschool – Second Edition (Wiig, Secord, & Semel, 2004). The CELF-P is an individually administered, norm-referenced measure of oral language ability. The core language subtests (Sentence Structure, Word Structure, and Expressive Vocabulary) assess knowledge of syntax, grammar, and expressive vocabulary. On the Sentence Structure subtest, participants are asked to point to one of four pictures that represents a target sentence. On the Word Structure subtest, participants are asked to complete a sentence using a targeted grammatical structure. On the Expressive Vocabulary subtest, participants are asked to verbally label a picture. The CELF-P has satisfactory evidence of validity and reliability (e.g., correlations with other measures of oral language, test–rest reliability of 0.77–0.92; Wiig et al., 2004). Scores on the three subtests are included in the core language score composite, which has a normative mean of 100 and a standard deviation of 15. The group mean on the CELF-P was 89.11 (SD = 8.98). On both the PPVT and CELF, standard scores were between 0.5 and 1.5 standard deviations below the normative mean. Scores in this range indicate poor oral language skills relative to a normative sample of same-age peers. Participant performances on identification measures are provided in Table 2.

All participants were African American and were age-eligible for kindergarten the following year. None of the participants had an identified disability or received services through an Individual Education Program (IEP). Demographic information was drawn from a family survey that was completed by parents or guardians of 17 of the 18 participants. Parental report on the survey indicated that all children and parents spoke English as their only language. Family income was reported for 15 participants, 12 of who reported income below the federal poverty line for their family size. Two parents reported receiving less than a high school education, 13 parents reported a high school diploma or GED, one parent reported some education after high school, and one parent reported a graduate degree.

Setting

The study was conducted at three public elementary schools with one pre-K classroom in each school. These pre-K classrooms served primarily children from families with low income. Many children in these classrooms had limited oral language skills; the group mean for all consented children who were administered the PPVT-IV (52 of 54 consented participants) was 90.82 (SD = 12.43). This mean is similar to group means of at-risk preschoolers from low-income families reported by other research groups ($M = 91$, SD = 11; Washington & Craig, 1999) as well as comparable to previous years of our own research in these classrooms. The classrooms were staffed by a lead teacher, one full-time assistant teacher, and one part-time assistant teacher. Programs were full-day classrooms with 18–20 children enrolled. Classrooms implemented the same comprehensive early childhood curriculum and reported some use of supplemental activities for children who were falling behind.

Intervention and measurement sessions took place outside the classroom in small tutoring or meeting rooms. Intervention sessions were conducted in small groups; all measures were administered individually to participants. When a room was not available, sessions took place in the hallway of the school. In all settings, efforts were made to reduce distractions and disruptions. For example, a session in the hallway would be located at the end of the hallway, away from bathrooms or other busy areas; tables and chairs would be positioned so that participants faced away from distraction.

Experimental design

To answer our research questions, we employed two experimental designs: a randomized control group design and an embedded single case design. In the randomized control group design, a total of 18 children, six in each of three classrooms, were randomly assigned to treatment or comparison (delayed treatment) conditions. In the treatment condition, participants completed a nine book series of storybooks with embedded lessons and three review books. Comparison participants received

Table 2
Characteristics of participants.

Treatment participants								
School	Child	Age	PN	WODB	PPVT-Pre	PPVT-Post	CELF-Pre	CELF-Post
	D1	4; 0	4	6	77	81	79	73
D	D2	4; 2	5	*	85	94	81	90
	D3	4; 11	6	4	84	88	79	88
	E1	4; 5	7	4	87	92	96	86
E	E2	4; 5	4	4	90	94	92	90
	E3	4; 10	6	9	78	97	96	98
	F1	4; 11	6	13	86	106	98	104
F	F2	4; 1	4	7	81	107	83	96
	F3	4; 11	8	*	83	88	98	94
Mean (SD)		4; 6	5.56 (1.42)	5.52 (4.15)	83.44 (4.22)	94.11 (8.42)	89.11 (8.43)	91.00 (8.75)
Comparison participants								
School	Child	Age	PN	WODB	PPVT-Pre	PPVT-Post	CELF-Pre	CELF-Post
	D4	4; 6	3	6	80	85	67	69
D	D5	4; 6	6	5	85	98	90	100
	D6	4; 1	3	*	80	90	86	94
	E4	4; 11	8	4	78	82	84	77
E	E5	4; 10	7	6	89	103	94	100
	E6	4; 7	5	6	88	105	96	100
	F4	4; 0	3	*	84	86	79	73
F	F5	4; 4	5	6	80	90	75	81
	F6	4; 1	4	6	87	92	84	102
Mean (SD)		4; 5	4.89 (1.83)	4.33 (2.55)	83.44 (4.07)	92.33 (8.05)	83.89 (9.21)	88.44 (13.31)

Note: Age at the beginning of the study is reported in years; months. PN = Picture Naming Individual Growth and Development Indicator (Bradfield et al., 2013). WODB = Which One Doesn't Belong Individual Growth and Development Indicator (Bradfield et al., 2013); * indicates that the child did not pass sample items and the test was discontinued. PPVT = Peabody Picture Vocabulary Test Fourth Edition (Dunn & Dunn, 2007), total standard score is reported. CELF = Clinical Evaluation of Language Fundamentals-Preschool (Wiig et al., 2004), total standard score for Core Language composite is reported.

'business-as-usual' classroom instruction. Treatment and comparison participants completed measures after each three-week unit.

Embedded in the group design, performance of the treatment participants was examined in a single-case experiment using a repeated-acquisition design. The repeated-acquisition design is an alternative to a multiple-baseline design when examining learning of multiple sets of target behaviors (Kennedy, 2005). The repeated-acquisition design allows for multiple demonstrations of learning across several sets of target behaviors to indicate that the intervention is responsible for learning. Experimental control is confirmed by consistent replication within participants (acquisition of instructional targets in nine books) and across participants ($n = 9$). In this study, treatment participants were exposed to a single intervention condition and learning of vocabulary and comprehension skills was examined within that condition. Thus, potential differences in intervention target difficulty were minimized as a threat to internal validity, because of the nine replications for each participant.

Procedures

The duration of the study, including pretesting, posttesting, and school breaks, was 14 weeks. In the group design experiment, treatment and comparison participants completed measures before and after each unit (a unit consisted of three instructional books and one review book). Thus, comparison participants were seen for individual testing sessions approximately every three weeks. In the embedded single-case design experiment, intervention sessions were conducted each school day as school schedules allowed. Treatment participants completed unit measures as well as mastery monitoring probes before and after each instructional book (approximately weekly). Thus, treatment participants were seen on an almost daily basis, for intervention or measurement sessions.

For treatment participants, the schedule for each instructional book was as follows: on the first day, participants listened to the plain storybook in small groups at the listening center and completed the mastery monitoring probe pretest individually. Next, participants listened to the instructional storybooks three times. Whenever possible, children listened to books on consecutive days. If a child was absent, the other two children listened to the book and the absent child received an individual "make-up" session as soon as possible. Participants completed the mastery monitoring probe posttest on the day after listening to the instructional book the third time. See Table 1 in online supplementary materials for sample intervention and assessment activities for one unit.

The intervention was administered to participants in the treatment group in small group 'listening centers' with three children and an adult facilitator. Facilitators, trained members of the research staff, were responsible for assisting students at the listening center (e.g., helping children stay on the correct page, keeping headphones on). Facilitators did not provide any instruction or modeling but frequently delivered praise or non-verbal approval to encourage participation. Facilitators completed daily checklists for procedural fidelity and attendance logs.

Review books were included after every three instructional books (approximately every three weeks). After completing the three listens each for the three instructional books, participants listened to the review book one time and completed the unit posttest.

Measures

Trained research staff administered all measures individually. In the group design, two assessments were used to measure primary outcomes: unit vocabulary tests and the *Assessment of Story Comprehension* (ASC; Spencer & Goldstein, 2011). The PPVT-IV and CELF-P were administered pre-intervention to identify and describe participants and post-intervention as measures of generalized intervention effects.

Unit tests

To assess knowledge of vocabulary, children were asked to provide a definition in response to an open-ended question (e.g., *What does [target vocabulary word] mean?*). We selected a definitional task to provide a rigorous measure of word knowledge, rather than a picture association or multiple-choice task in which chance level responding might overestimate what children know. Definitional tasks have the potential to be sensitive to incremental development of word knowledge (Beck et al., 2013) especially when a scaled scoring system is used. Responses were scored on a scale: zero points for an incorrect response; one point for a related or partial correct response; and two points for a complete and correct response. Cronbach's alpha internal consistency coefficient for this sample was 0.86. For newly developed measures such as the unit test, Gersten et al. (2005) suggest that acceptable levels of internal consistency are 0.60 or greater.

Unit tests were administered prior to and immediately following the completion of each unit (each unit was comprised of three instructional books and one review book). Thus, for each unit, participants had a pretest and posttest score. The unit test was most often administered on the same day that participants listened to the review book.

Assessment of story comprehension

The Assessment of Story Comprehension (ASC) was used to measure improvements in question answering. The ASC is a researcher-made curriculum-based measurement tool with nine parallel forms (i.e., nine stories). Preliminary investigations indicate that the ASC has satisfactory psychometric properties (Spencer, Goldstein, Kelley, Sherman, & McCune, 2015). The ASC has strong correlations with other measures of oral language (e.g., $r = 0.81$ with CELF-P), adequate parallel form reliability (0.65–0.83), and high internal consistency (Cronbach's alpha = 0.96).

Following standardized, scripted administration procedures, an examiner reads one of the stories and asks a mix of literal and inferential questions about the story. Three questions are literal questions about the key story events (e.g., *What was Jenny doing in this story?*). The remaining four questions closely align with the types of questions addressed in the Story Friends lessons: questions about character emotions (e.g., *Why was Jenny sad?*), character actions (e.g., *Jenny's teddy bear fell in the mud. What happened next?)*, post-story predictions (e.g., *Next time Jenny plays outside, do you think she'll take her teddy bear? Why/why not?)*, and questions that require children to use background information (*Why do you think Jim wanted to help Jenny?*). Administration takes approximately 5 min.

The ASC includes standardized scoring procedures using scoring guides with story-specific sample responses. For items 1–7, responses to questions are scored for clarity and completeness on a three-point scale. For example, if a child's response to the question, "Why was Jenny sad?" was correct, complete, and clear (e.g., "Her bear fell in the mud.") it was given a score of two, and if the response was correct, but incomplete or unclear (e.g., "It's dirty.") it was given a score of one. If the response was incorrect (e.g., "Playing in the mud.") it was given a score of zero. The final item addresses a less common word that is contextually supported in the story. It is a two-part item, with the initial question asking for an expressive definition of the word (e.g., "Tell me, what does *filthy* mean?"). If the child's response is fully accurate, a synonym, or provides definition in an example it is given a score of three and administration ends. If the response is not an obvious 3-point response, the examiner immediately asks a follow up question (e.g., "Does *filthy* mean very tall or very dirty?"). The child's responses to both parts are recorded on the protocol so an accurate score can be determined later, if necessary. If the response to the first part is unclear but related to the story or an example of the word without a definition, it is given a score of two, even if the second part had to be administered. However, if the response to the first part did not earn two points, then the response to the second part of the question is considered. One point is given if the child responds with the correct choice of the two options (e.g., "very dirty"). The ASC has a total possible score of 17 with seven questions scored on a 0–2 scale and the one two-part question scored on a 0–3 scale.

The ASC was administered immediately following the completion of the unit tests. Random sequences of ASC stories were generated for each participant so that they received the parallel forms in a different order. Importantly, participants had never heard ASC stories and questions during intervention. ASC stories were completely unfamiliar to the children and because of their novelty the ASC offers a more distal and conservative measure of intervention effects than the mastery monitoring probes.

Mastery monitoring probes

For the repeated-acquisition design, treatment participants completed mastery monitoring probes to assess learning of instructional content in each book. Mastery monitoring probes were researcher-created measures of the instructional content in each book. The mastery monitoring probes for Books one through six consisted of five items: two that assessed learning of vocabulary and three that measured responses to story questions. Vocabulary items were identical to the unit tests (e.g., *what does [target word] mean?*). Story question items were exact repetitions of the story questions included in the embedded lessons. All items were scored on a scale of 0, 1, 2: 0 points for an incorrect response; 1 point for a related or partial correct response; 2 points for a complete and correct response.

The mastery monitoring probes for Books 7, 8, and 9 included additional vocabulary items for untaught words. The untaught words were included to examine the possibility that participants might learn vocabulary through simple exposure to the words through the repeated readings or assessments. Untaught words were included in the story text but embedded lessons were not provided for the untaught words. To be appropriate comparisons for target vocabulary, these words were selected to be similar in difficulty (e.g., represent Tier-2 words, verbs rather than nouns), occurred with the same or greater frequency in the story text, and to have some support from story context or illustrations. For example, the untaught word for *Ellie's First Day* was *giggle*. This word was a more sophisticated synonym for a familiar word (e.g., *laugh*) and occurred in a supportive context of the story (e.g.," All the kids started giggling, 'Ha ha hee hee!' 'Oh no!' thought Ellie. 'They are laughing at me!'" on a page with an illustration of story characters laughing.) Mastery monitoring probes were administered at pre-book and post-book measurement points for each of the nine instructional books for treatment participants only; they were not administered to the comparison group so as to not frustrate children by asking them weekly to respond to items not taught. During pre-book test sessions, participants listened as a group to the 'plain' versions of the stories. Then, participants were assessed individually using the mastery monitoring probes. The 'plain' book was included to allow for fair measurement of treatment effects. It would have been inappropriate to ask participants to respond to questions about a story they had not yet heard. By listening to the 'plain' story, participant responses indicated what children knew prior to receiving the embedded lessons.

Fidelity of implementation and assessment

Trained observers monitored implementation fidelity for listening centers. Observers scored a fidelity checklist of seven items related to key components of the listening center and to facilitator behavior (each child had a book, all children were wearing

headphones, facilitator was wearing headphones, correct, complete audio was played, children were reminded of behavior expectations, facilitator provided non-specific positive feedback, facilitator did not provide any additional instruction). Observations were conducted in-person or by coding videotapes of listening centers and the facilitators were aware that the session was being observed. When observations were conducted in-person, the observer sat away from the participants to remain unobtrusive. Implementation fidelity observations were distributed across schools and facilitators and balanced across books and the first, second, and third listen. Implementation fidelity assessed for 26% of 90 sessions was scored at 100%, indicating that key components of the listening center were always in place.

Fidelity of implementation also was examined by reviewing attendance log records of the dates that children listened to storybooks. The goal was for children to hear each book without embedded lessons one time, each book with embedded lessons three times, and each review book one time. All participants received this intended dosage. A secondary goal was that children listened to the books with embedded lessons on consecutive days (e.g., 2 days between the first and third listen). Across participants, the average amount of time between the first and third listens was 3.8 days. Weekends, school breaks, and absences contributed to the extended time between listens.

To examine fidelity of test administration, trained research assistants listened to audio recordings and completed procedural checklists specific to each measure. Fidelity checks were distributed across pre- and post-book assessment times and across examiners and participants. Fidelity was examined separately for mastery monitoring probes, unit tests, and the ASC.

For the mastery monitoring probes, administration fidelity was examined for 32% of 162 assessment sessions. For vocabulary, mean administration fidelity was 99% (range 83–100%) for pre-book assessment and 99% (range 83–100%) for post-book assessment. For story questions, mean administration fidelity was 92% (range 50–100%) for pre-book assessment and 94% (range 83–100%) for post-book assessment. For vocabulary unit test, administration fidelity was collected for 23% of 108 assessment sessions. Mean administration fidelity was 97% (range 83–100%) at pretest and 99% (range 94–100%) at posttest. For the ASC, fidelity of administration was examined for 55% of 72 assessment sessions. Mean administration fidelity for the ASC was 99% (range 92–100%). Across all measures, the most common error was that the examiner did not deliver a prompt when it was warranted.

Scoring reliability

A trained research team member served as the primary scorer for all measures. An independent research assistant scored one third of the all the assessments to evaluate scoring reliability. Scoring was completed using detailed scoring guides created for each measure. For example, unit test scoring guides included a scoring rubric as well as multiple sample responses for each item. Scoring reliability was evaluated periodically during the study. Scorers were blinded to participant and assessment points (e.g., pretest/posttest). On the unit test and the ASC, scorers were blind to condition (treatment vs. comparison group) as well. It was not possible for scorers to be blind to condition on the mastery monitors because only treatment participants completed them. On all measures, an item-by-item comparison was made to determine agreement or disagreement. Scoring reliability was calculated by dividing the total number of agreements by the total number of agreements plus disagreements and multiplied by 100.

On the ASC, scoring reliability was calculated for each assessment point and ranged from 83 to 94% with an average of 90%. For the unit tests and mastery monitoring probes, the high number of 0 point answers (e.g., *I don't know* or no response) at pretest made scoring agreement more likely. Thus, we examined agreement separately for pretest and posttest. On the unit test, mean agreement was 98% at pretest (range 94–100%) and 99% at posttest (range 94–100%). On the mastery monitoring probes, agreement for vocabulary items was 100% across all pre-book and post-book assessments. For question-answering items, mean agreement was 96% across pre-book assessments (range 78–100%) and 88% across post-book assessments (range 67–100%). Although scoring reliability overall was quite high, the open-ended response on the question answering response posed a particular challenge for scoring agreement. All disagreements were between 2 point and 1 point scores, indicating that discriminating between partially correct and complete and correct responses was difficult.

Results

Group design results

To examine the equivalency of the two groups (treatment, comparison) at pretest, comparisons were conducted for standard scores on the PPVT-IV and CELF-P and for chronological age. There were no significant differences between groups or between classrooms on these variables.

Vocabulary

Table 3 presents the average gain scores for each unit test for each group. We conducted group comparisons using analysis of variance on the gain scores for each of the three unit tests using Bonferroni adjusted alpha levels of 0.016. Results indicated that participants in the treatment group had greater gains than comparison participants for each of the three units [$F(1, 17) = 24.33, 8.53, 30.86$ for Units 1, 2, 3 respectively, $p \leq 0.01$]. Effect sizes, Cohen's d, calculated using the gain scores for each unit test were large: Unit 1, $d = 2.33$; Unit 2, $d = 1.37$; Unit 3, $d = 2.62$.

Pretest scores on the unit tests were generally low for both groups ($M = 0.70$, SD = 1.11) indicating that children had very little knowledge of the vocabulary words prior to instruction. Rarely, children were able to provide a partial knowledge response (11 1-point responses of the possible 324) or a full knowledge response (13 2-point responses of 324). Significant gains were evident for the Treatment group for each unit test with post-test scores averaging 6.22 of a possible 12 points (SD = 3.72), whereas post-test scores for the comparison group averaged 0.57 (SD = 0.93).

Comprehension

For the ASC total score, there was no significant effect for group, a significant effect of time [$F(3, 48) = 4.05$, $p = 0.01$, $\eta_\rho^2 = 0.20$], and no significant group × time interaction. Because the Story Friends program specifically targeted inferential questions, we hypothesized that group differences would emerge on the inferential questions. To test this hypothesis, two subscores were calculated: performance on the three literal questions (maximum of 6 points) and performance on the four inferential questions (maximum of 8 points). For each subscore, a mixed 2×4 analysis of variance was conducted with one between-subjects factor (Group) and one repeated-measure factor (subscore score on the four ASC assessments). For the literal questions, there was no main effect of group or time and no group by time interaction. For inferential questions, there was no main effect of group or time, but a significant group × time interaction [$F(3, 48) = 4.86$, $p \leq 0.01$, $\eta_\rho^2 = 0.23$]. Group means for the ASC scores across time points are presented in Table 3 and subscores are graphically displayed in Fig. 1.

Table 3
Group performance on unit tests and the assessment of story comprehension.

Measure	Treatment ($n = 9$)		Comparison ($n = 9$)		ANOVA F		
	M	SD	M	SD	Group (G)	Time (T)	$G \times T$
Unit test gain							
Unit 1	5.00	2.69	0.22	1.09	24.33*		
Unit 2	4.44	4.67	−0.11	0.33	8.53*		
Unit 3	6.33	3.04	0.33	1.12	30.86*		
ASC total					1.63	4.05*	1.96
Pretest	5.22	3.70	4.89	4.51			
Time 2	6.89	2.20	6.33	4.70			
Time 3	8.11	3.55	4.56	3.13			
Posttest	9.56	4.01	6.44	4.19			
ASC literal					2.16	1.69	1.50
Pretest	1.22	2.05	1.11	1.97			
Time 2	2.56	1.67	1.56	1.88			
Time 3	2.89	1.90	0.89	0.93			
Posttest	2.33	1.5	2.00	2.06			
ASC inferential					0.992	2.14	4.86*
Pretest	2.56	1.67	2.56	2.30			
Time 2	2.67	1.41	3.11	2.32			
Time 3	3.89	2.03	2.89	2.26			
Posttest	4.78	1.72	2.22	2.05			

Note: Unit tests include items for the 6 vocabulary targets included in 3 books. The maximum score was 12 (2 words per book, 2 points per word). ASC = Assessment of Story Comprehension. On the ASC, the maximum score was 17 for total, 6 for literal questions, and 8 for inferential questions.
* $p \leq 0.01$.

Generalized effects

To examine generalized effects of the intervention on language skills, a mixed 2 × 2 analysis of variance was conducted with one between-subjects factor (group) and one repeated-measure factor (pretest, posttest) on standard scores of the PPVT-IV and CELF-P2. On the PPVT-IV, there was no significant effect for group, $F(1, 16) = 0.12$, $p = 0.74$. There was a significant main effect of time, $F(1, 16) = 34.21$, $p \leq 0.001$, $\eta_\rho^2 = 0.68$, but no significant group × time interaction, $F(1, 17) = 0.28$, $p = 0.60$. At posttest, the effect size of the difference between the treatment and comparison groups was small ($d = 0.22$). On the CELF-P2, there was no significant effect for group, $F(1, 16) = 0.78$, $p = 0.39$, or time, $F(1, 16) = 3.07$, $p = 0.10$, and no significant treatment-by-time interaction, $F(1, 16) = 0.53$, $p = 0.48$. At posttest, the effect size of the difference between the treatment and comparison groups was small ($d = 0.23$). Group means are presented in Table 2.

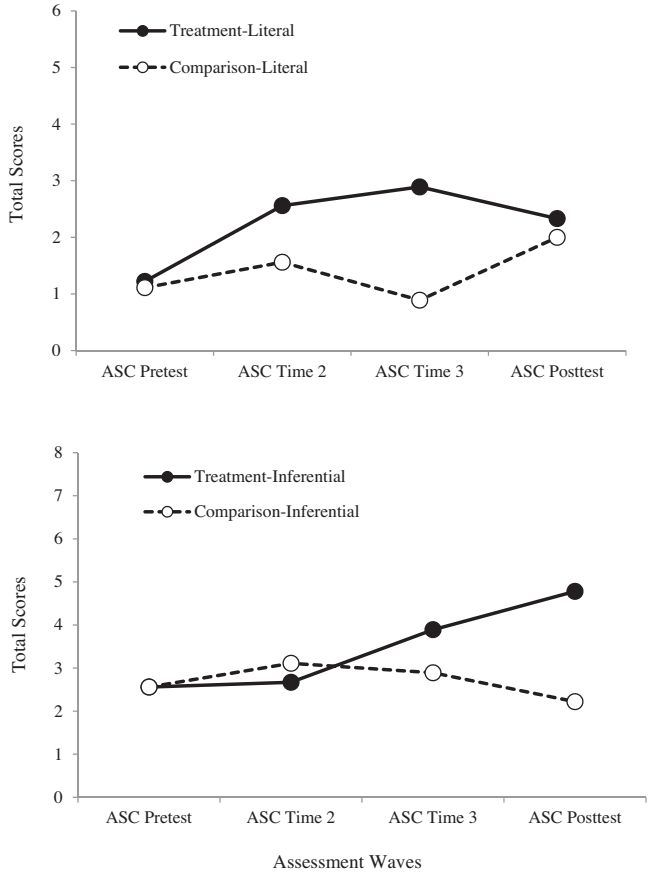

Fig. 1. Mean scores for literal and inferential questions on the Assessment of Story Comprehension (ASC) for participants in the treatment and comparison groups. Maximum score was 6 for literal questions and 8 for inferential questions.

Repeated-acquisition design results

In single-case experimental design experiments, evidence of treatment or experimental effect is shown when predicted changes in outcomes occur when the intervention is implemented (Horner et al., 2005). In this study, we predicted an increase in performance on measures of vocabulary and comprehension after children received the embedded storybook intervention. To identify evidence of treatment effects within the single-case experimental design, figures were created to include a panel for each child's vocabulary and comprehension performance separately (see Figs. 2 and 3). We set criteria of two points for a treatment effect for each domain: vocabulary and comprehension. For each child, the pre-book score and post-book score for each book were plotted. An increase between pre-book and post-book of two points was evidence of a treatment effect. For example, in Fig. 2, for book 4 vocabulary, child D1 had a vocabulary score of zero at pre-book (open circle) and two at post-book (closed circle). For each participant, nine replications of treatment effects (one per book) were possible for vocabulary and for comprehension. For example, in Fig. 2, child D1 had 5 replications of treatment effects (books 4, 6, 7, 8, 9). Across the nine participants, there were 81 possible replications of treatment effect for each domain.

Vocabulary

For the vocabulary outcome, a treatment effect was defined as a 2-point difference between pre-book and post-book measurement. A 2-point increase could be achieved by an improvement

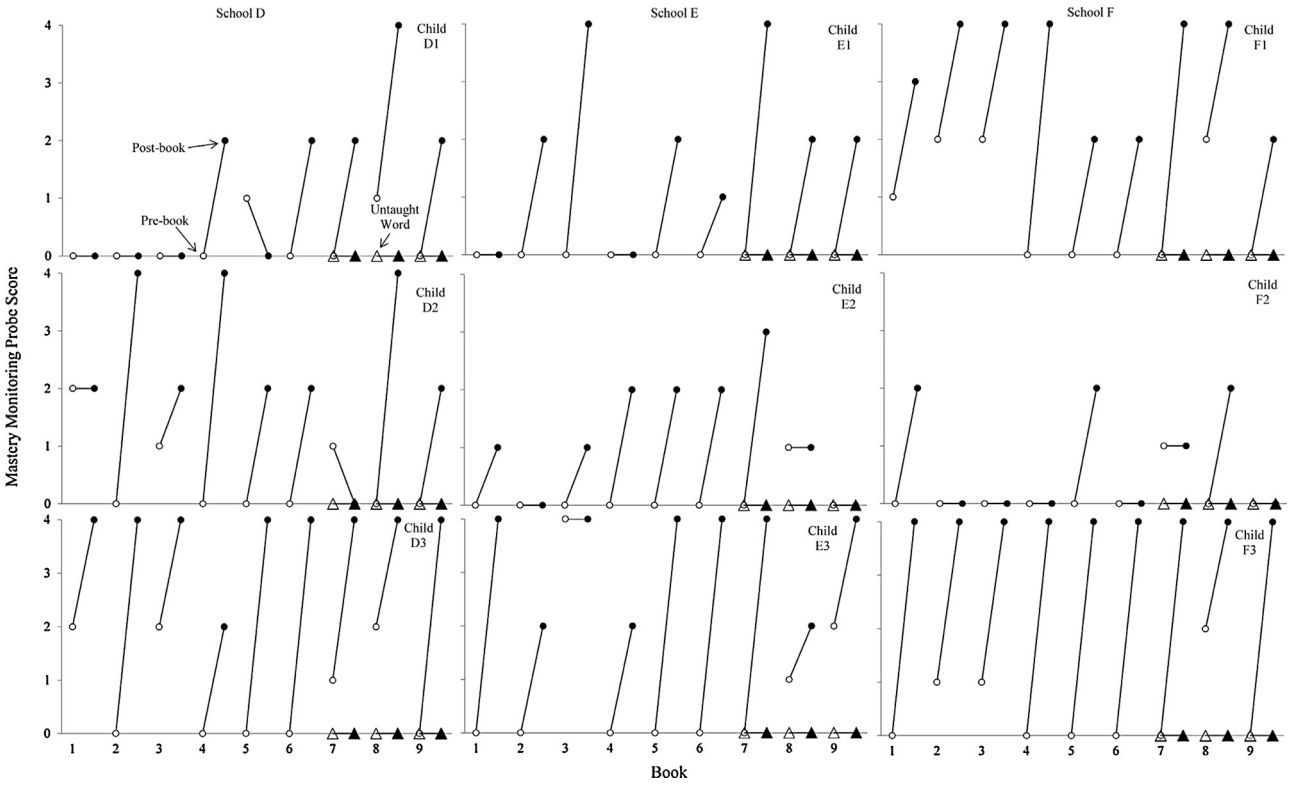

Fig. 2. Vocabulary scores on the mastery monitoring probes for participants in the treatment group. Maximum score for the pre-book (open circles) and post-book (closed circles) assessments was 4 for the two vocabulary targets for each book. Untaught words denoted by triangles were assessed for three books. For untaught words, the maximum score was 2. Improvements from pre-book to post-book can be evaluated for nine books for each of nine participants (i.e., 81 possible replications).

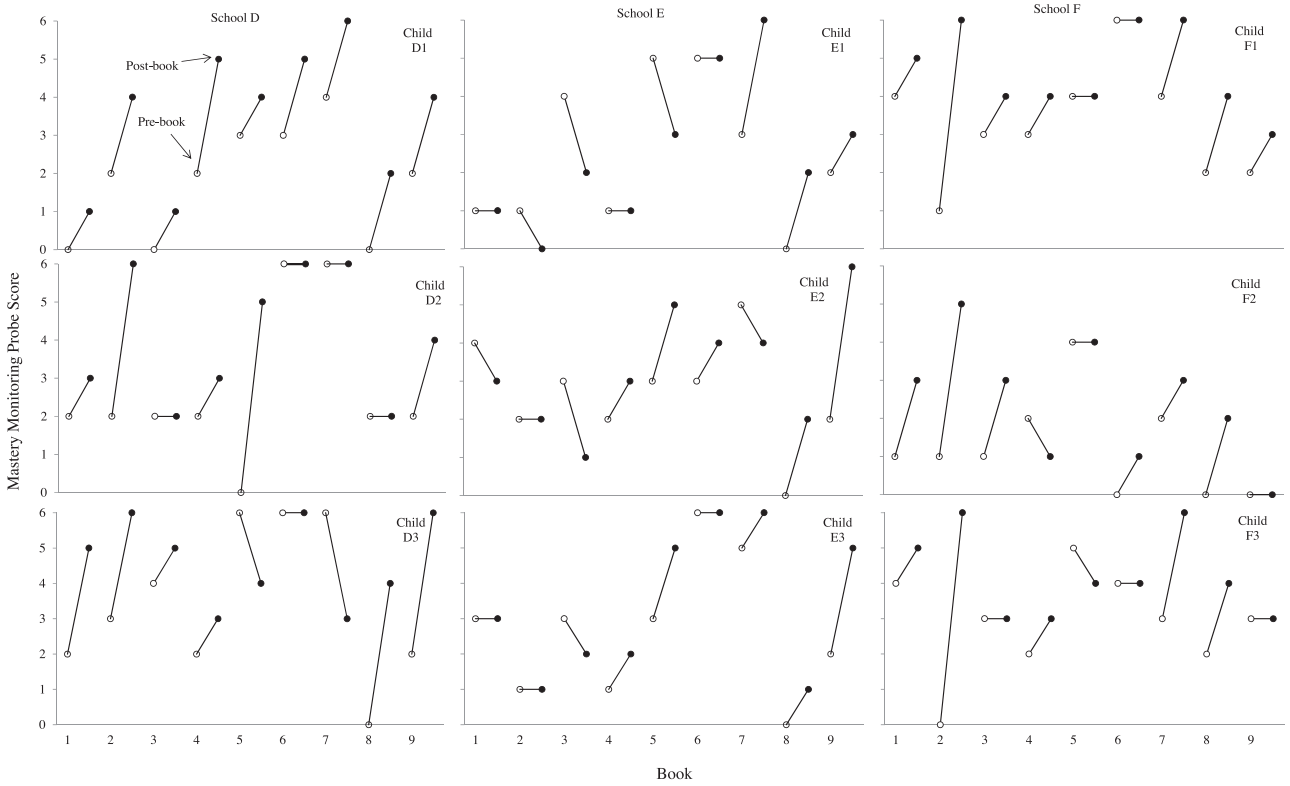

Fig. 3. Comprehension mastery monitoring probe scores for participants in the Treatment group. Pre-book (open circles) and post-book (closed circles) assessments had a maximum score of 6 for the three comprehension questions for each book.

on a single word (e.g., an increase from incorrect to correct) or by improvements on two words (e.g., increases from incorrect to partially correct or partially correct to correct). Treatment effects were evident for at least some books for all participants. The mean number of books in which participants showed a 2-point difference was 6.44, (range 3–9). Given the 81 possible demonstrations of treatment effects, there were 58 replications of treatment effects on the vocabulary mastery monitoring probes. To examine word level results, gain scores were calculated for each child for each word. On average, children learned 10 words (range 3–18). Fig. 2 includes the vocabulary results for all children. There was substantial variability in the number of words children learned; child F2 was able to provide just 3 definitions whereas child F3 provided 18 definitions. As expected, participants demonstrated no knowledge of untaught words on pre- or post-book assessments.

Comprehension

For the comprehension outcome, a treatment effect was defined as a 2-point increase from pretest to posttest. A 2-point increase could be achieved by an improvement on a single question (e.g., an increase from incorrect to correct) or by improvements on two questions (e.g., increases from incorrect to partially correct or partially correct to correct). Similar to the vocabulary results, all children showed a treatment effect for some of the books, although the mean number of books with effects was lower ($M = 3.33$, range 3–6). Pre-book scores were higher than for the vocabulary domain ($M = 2.55$, SD = 1.75), indicating that participants could frequently answer questions prior to intervention. On average, post-book scores were 1.11 points higher than pre-book scores (SD = 0.82). Across participants, there were a total of 30 replications of treatment effects out of the 81 possible demonstrations.

Discussion

Treatment outcomes on proximal and distal measures of language

Results from both the group design and single case design studies indicated that an automated Tier-2 intervention was generally effective in teaching challenging vocabulary words. In the group design study, effect sizes were large for targeted vocabulary, indicating that children who participated in the intervention learned challenging vocabulary words (combined unit test average of 18.7 of a possible 36 points; 52% correct) whereas children who received 'business-as-usual' classroom instruction did not (combined unit test average of 1.78 points; 5% correct). This result extends the substantial body of evidence that explicit vocabulary instruction is important for vocabulary learning (Beck & McKeown, 2007; Coyne et al., 2007; Loftus et al., 2010) and that the quality of oral language instruction in many classrooms may be limited (Dickinson, 2011; Justice, Mashburn, Hamre, et al., 2008; Justice, Mashburn, Pence, et al., 2008). Explicit vocabulary instruction occurs infrequently in classrooms (Wright, 2012); it is unlikely that children in the comparison group were exposed to explicit vocabulary instruction on the targeted words.

In the single-case design, treatment participants learned an average 10 of a possible 18 challenging vocabulary words (range 3–17 words). Children did not learn any of the untaught vocabulary words, indicating that learning was likely resulted from the embedded lessons rather than exposure to the words via repeated readings or testing. We explored factors that may have contributed to different responses to intervention by examining correlations between number of words learned with age, prior vocabulary knowledge, and oral language skill. Only age emerged as a significant predictor of word learning ($r = 0.85$, $p = 0.004$); children who were older learned more words than children who were younger. Given the narrow range of age of participants (4;0–4;11), this finding is unexpected. A frequent finding in other vocabulary intervention studies has been a correlation between prior vocabulary knowledge, often scores on the PPVT, and a response to vocabulary instruction (Coyne et al., 2007, 2009; Loftus et al., 2010). In this group of participants, all of whom had below-average scores on the PPVT-IV (standard scores between 77 and 90), we did not detect a correlation.

In the group design study, performance on the ASC indicated that the intervention improved the participants' question-answering skills with respect to inferential questions, but not for literal questions. Effect size at posttest for inferential questions was large ($d = 1.10$); children in the treatment group performed substantially better than children in the comparison group on these types of questions. Although the intervention did not impact literal question-answering skills, the observed effect on inferential questions is noteworthy. The content of the ASC stories and inferential questions, although similar to the type of questions addressed in intervention, were totally new, which provided a stringent measure of effects.

In the single-case design, the intervention was less impressive for the comprehension outcome than for the vocabulary outcome; treatment effects were observed for an average of only 3.33 books per child. The relatively low dose of intervention may be responsible for limited effects in the single case design. For each question, children received just three opportunities to respond and to hear the modeled responses. In contrast, children received many more opportunities to respond to each vocabulary word (two lessons per word in each book with multiple opportunities to respond for each lesson). The limited instruction on individual story comprehension may have contributed to modest effects on the mastery monitoring probes; however, over the course of the nine books, the impact on inferential question-answering was favorable, measured using the ASC. Although stronger effects are typically seen on proximal measures than distal measures, in this case, the more conservative measure of question-answering ability yielded stronger effects.

Although these findings indicate that although this intervention has promise for improving question-answering abilities, there is room for improvement. Because the teaching procedures were embedded in automated storybooks and did not involve contingent feedback and obligatory responding, children received a fairly low dose of comprehension instruction. To substantially improve question-answering skills, increased dosage with more opportunities to respond may be necessary. Results from the ASC also indicate that intervention efforts should be aimed at both literal and inferential questions. Performance on the literal questions on the ASC was quite low (average score of 1.81 of a possible 6). Although our previous development work indicated that many children could answer literal questions, findings from this study suggest that there is a need for instruction that targets literal questions about the story (e.g., *What was Danny doing in this story? What happened at the end of the story?*). Gains in inferential questions were observed; however, treatment participants did not reach mastery levels (average posttest score of 4.78 of a possible 8). Thus, additional intervention efforts to target inferential questions also are warranted. Further research is necessary to identify the specific questions that would be successful targets for this type of intervention. However, the inferential question types included in this intervention are a promising start.

One challenge in the intervention design was to identify story questions that children could not already answer prior to intervention. Because children were able to successfully answer inferential questions prior to intervention fairly frequently (average pretest scores on the mastery monitor scores of 2.56 of a possible 6, SD = 1.75), there was less room for improvement. We carefully

examined performance on story questions in an attempt to identify questions that were the most successful intervention targets. It is likely that the restricted scoring range (0–2 points possible) may have limited our ability to detect subtle changes in performance on the question types. Unfortunately, no distinct patterns of performance emerged. Pretest performance did not differ substantially across the different question types and pre-post differences were similar within question types. There may have been a slight advantage at pretest for inferences about character emotions (pretest average of 1.05 of a possible 2 points) relative to other question types (pretest averages of 0.91 for inferences about character actions, 0.82 for post-story predictions, 0.62 for connections with children's lives) which meant that gains on character emotion questions were smaller.

Although efforts were made (e.g., restriction of number of question types) to create intervention questions that were somewhat equivalent, it is clear that questions ranged in difficulty. For example, at pretest, the question "*Why did Leo tell Marquez to turn around?*" had just one correct 2-point response whereas the question of "*Why did Leo's friends come to see him?*" had 8 of 9 treatment participants providing a correct 2-point response at pretest. Both questions targeted an inference about character action and referenced key story events, but perhaps the latter may be easier to guess based on prior experience. Observations by research staff indicate that children were more interested in particular stories or characters, which also may have contributed to differences in performance.

No significant effects of treatment were observed on standardized norm-referenced measures of language ability, the PPVT-IV and the CELF-P. Few studies of supplemental, targeted interventions with young children have provided evidence of effects on such distal measures. Many intervention studies report findings on intervention-specific measures (e.g., Beck & McKeown, 2007; Coyne et al., 2007; Justice et al., 2005; Loftus et al., 2010); general outcome measures, such as standardized, norm-referenced measures have been included less frequently. Coyne et al. (2010), examined transfer effects of a supplemental vocabulary intervention and reported a moderate ($d = 60$) but not statistically significant effect on the PPVT. The multiple regression analysis revealed a small effect size of 0.14 associated with a standard score of 85 on the PPVT. In a meta-analysis of vocabulary interventions in prekindergarten through 12th grade, Elleman et al. (2009), reported an effect size of 0.29 on norm-referenced measures of vocabulary. van Kleeck and colleagues reported large treatment effects of an intensive, individual question-answering intervention on the PPVT (omega squared of 0.16). In the current study, the magnitude of the difference between treatment and control groups at posttest was 0.22 on the PPVT, an effect size consistent with those reported by Elleman and Coyne and their colleagues. Participants in both treatment and comparison groups made significant gains on the PPVT during the intervention period (pre-post gains of $d = 1.53$); this finding might be attributed to Tier 1 instruction in these classrooms.

Implications of treatment outcomes

The current study presents evidence of a promising supplemental intervention to be provided as part of an effective RtI model. Given the limited oral language skill of participants prior to intervention, the challenging nature of the vocabulary words, and the rigorous definitional measurement task, we are encouraged by these findings. If incorporated into an RtI model, in which this intervention would be provided as a supplement to a high-quality, language rich Tier-1 curriculum, decontextualized knowledge of 10 of 18 challenging vocabulary words compares favorably to other vocabulary interventions.

Although vocabulary intervention studies frequently include estimates of effect sizes, researchers have rarely reported findings in terms of number of words learned. However, we can estimate the number of words from reported scores on researcher created measures. To consider the impact of an intervention, it is valuable to consider the amount of instruction as well as resulting gains in knowledge. In the current study, participants completed a total of 7.5 h of instruction and learned an average of 10 words. This proportion (56% of taught words) is larger than what has been found in previous vocabulary intervention studies (e.g., 27% of targeted vocabulary in Penno et al. (2002), less than 10% in Justice et al. (2005). and translates into 45 min of instruction per word learned. Coyne et al. (2010) reported average gains of 50.15 points on the target word measures. With a possible four points per word, a gain of 50 points might roughly indicate learning of 12 words. Participants received 18 h of instruction (90 min of instruction per word learned). Using an expressive definition task similar to the one included in this study (2 points maximum per word), Loftus et al. (2010) reported knowledge of slightly more than 1 word (2.2. points) following 4 h of instruction (240 min per word). Pullen et al. reported knowledge of 3–4 words (3.67 points, 1 point per word) following 200 min of instruction (~55 min per word). Justice et al. (2005) provided 400 min of instruction; participants gained 2.3 words (~175 min per word). It appears that time-intensive instruction is necessary to produce changes in decontextualized vocabulary knowledge of at-risk children and that, in comparison to other supplemental vocabulary interventions, participants in *Story Friends* a greater proportion of taught words with less instructional time. Nevertheless, future research will need to determine if more robust effects can be achieved if, for example, teachers reinforce automated vocabulary intervention through additional classroom instruction.

Fidelity of implementation

As treatment research moves into field-based settings, implementation fidelity presents a substantial challenge. It can be difficult for teachers to implement interventions with the level of precision intended by the developers of a program. For example, Dickinson (2011) reported fidelity of just 38% of instructional elements in shared book reading. Language facilitation strategies may be particularly difficult for teachers to implement (Pence, Justice, & Wiggins, 2008; Piasta et al., 2012). Even with intensive professional development and coaching, teachers rarely include decontextualized talk about vocabulary during shared book reading (Wasik & Hindman, 2014).

We designed the *Story Friends* intervention using an innovative automated approach with the goal of delivering evidence-based vocabulary and comprehension instruction with high fidelity in authentic educational settings. In several ways, we were successful in achieving this goal. All participants received the intended dosage of explicit instruction (e.g., listened to each instructional book three times). Because instruction was delivered via prerecorded audio, rates of explicit instruction, decontextualized talk about vocabulary (e.g., connections to children's lives), and inferential questions were high. Findings of the current study suggest that the intervention could be implemented with comparable fidelity by educational staff, as the intervention required little training of staff. Facilitators were trained to provide only minimal input (e.g., encouragement to stay on task, help turning pages); facilitators provided no additional instruction or corrective feedback. We might expect similar interactions from educational staff (e.g., instructional aides) who receive no specific training in explicit vocabulary intervention.

Limitations and future directions

There are a number of limitations worth noting, many of which relate to future research and continued iterative research and development of *Story Friends*. A major limitation of the current study is the small sample size for a group design study. The small sample size did not allow for consideration of classroom-level variables that may contribute to intervention effects. Because the intervention was delivered in small groups, it is also possible that group peer effects may have contributed to performance differences. In future studies, larger sample sizes will be used to provide additional information about the source of intervention effects. Larger samples would provide evidence that may be more readily generalizable, as well as allow for analyses to identify child- and classroom-level predictors of success.

We did not conduct observations of classroom instruction during the study making it difficult to characterize the 'business-as-usual" experiences of participants in the comparison group. However, our previous work in these classrooms, as well as reports by other research groups, suggests that instructional support for language is limited (Dickinson, 2011; Greenwood et al., 2012; Justice, Mashburn, Hamre, et al., 2008; Justice, Mashburn, Pence, et al., 2008). For children in early childhood programs with limited oral language skills, the Tier-1 instruction provided in typical classrooms will be unlikely to close skill gaps (Greenwood et al., 2012). Without convincing evidence of high quality instruction at Tier-1, it is difficult to examine the effects of a Tier-2 supplemental intervention. In future studies, comparisons with alternative treatments, for example teacher-led small-group language intervention may be warranted.

Future work also must address the issue of long-term and generalized effects of targeted interventions. An important element of long-term effects is retention of the word knowledge acquired in the intervention. We were able to collect retention data 6–7 weeks later, but only for the Unit 3 vocabulary items. Participants were only assessed on vocabulary items for which they produced 2-point definitions. Thus, the number of words assessed varied by participant (e.g., Child F2 was assessed on only the 3 words he learned). Across the nine participants, word knowledge was maintained for 47% of previously learned words (16 of 34 words assessed); word knowledge was reduced from complete knowledge (score of 2) to partial knowledge (score of 1) for 12% (4 of 34 words); children did not maintain word knowledge for 38% (13 of 34 words). Other researchers who have examined maintenance of word knowledge have reported mixed findings. In some cases, word knowledge has been maintained (Loftus et al., 2010); in others, word knowledge has increased between posttest and the follow up 6 weeks later (Biemiller & Boote, 2006). Children with limited vocabulary knowledge prior to intervention may be less likely to maintain newly acquired word knowledge. Pullen et al. (2010) reported that on a probe for maintenance four weeks after a 2-week intervention, children with low scores on a vocabulary measure demonstrated a decrease in word knowledge from posttest scores whereas children with higher vocabulary scores were able to maintain their newly acquired word knowledge. Given the challenging words taught in this intervention, and the limited vocabulary knowledge of participants prior to intervention, we are encouraged that word knowledge was maintained for half of the learned words. However, the small amount of retention data available limits the extent to which the intervention can be evaluated for long-term effects.

To promote long-term maintenance, growth of word knowledge, and generalized effects on language skills, it is apparent that children will require learning opportunities that extend beyond this brief intervention. No treatment differences were observed on the two general outcome measures of language, suggesting that stronger or different interventions are necessary to produce changes on these types of measures. It is notable that the words targeted do not appear on the PPVT unless they appear much later in the test. Future research should investigate other potential long-term outcomes. Vocabulary intervention has been demonstrated to improve passage comprehension of older children (Clarke, Snowling, Truelove, & Hulme, 2010); similar improvements may occur in listening comprehension with young children. Metalinguistic knowledge or other word learning measures also might be sensitive to changes produced by this intervention. As children become more experienced in word learning, perhaps they can become better word learners and become more efficient in acquiring new words from linguistic input.

This revision of *Story Friends* appears to represent an improvement on previous versions (e.g., average learning of 10 words in the current study vs. 8 in Spencer et al., 2013) and presents an advantage relative to many other vocabulary intervention studies in terms of instructional time and proportion of words learned. However, there remains room for improvement. Children learned only a portion of the words taught, and the question remains whether learning of a small number of words will have long-term effects on later reading acquisition.

The automated format of the intervention presented some challenges that might be addressed by live delivery by trained educational staff. The prerecorded lessons are consistent regardless of participant behavior (e.g., modeled responses were the same regardless of children's responses). For the story questions in particular, this was problematic. Stronger student effects may be observed when teachers are available to provide contingent feedback and scaffold student responses. A combination of teacher-led and automated approaches might be effective. Teacher-led interactive activities using digital media have been used to improve the early literacy skills of preschool children (Penuel et al., 2012). Intervention procedures were consistent across children; children listened to instructional books three times regardless of their mastery of, or lack of mastery of, instructional content. It may be important to examine ways to modify the intervention for individual children. One particularly intriguing option would be the use of a computer-based format. A computer-based format would allow for adaptation of delivery, timing, and feedback to individual children's responses and increase the interactive options used to engage the children.

Given the high number of children with limited oral language skills in many early childhood settings, this intervention might be an appropriate activity for all children, perhaps during center or choice times. This might be particularly feasible if children could participate in *Story Friends* activities without adult support. In the current study, we observed that many children required little support from the facilitator; it is possible that with clear pre-teaching of expectations (e.g. keeping headphones on), little practice, and minimal supervision, small groups of children could rotate through the listening center. Whole group delivery might also be effective for supplemental vocabulary instruction (Neuman & Kaefer, 2013).

Conclusion

The purpose of this study was to examine the effects of an automated Tier-2 intervention to improve the vocabulary and comprehension skills of prekindergarten children with limited oral language skills. Findings indicate that the intervention resulted in improved word knowledge and moderate effects on comprehension skills. Based on these findings and the limitations of this study, further examination of *Story Friends* is warranted using a larger scale efficacy trial. The intervention, utilizing automated storybooks and embedded lessons, appears feasible for implementation

Table A1
Books, targeted vocabulary, and untaught words.

Book	Targeted vocabulary	Untaught words
Ellie's First Day	enormous, different	giggle
Leo's Brave Face	brave, grin	arrive
Jungle Friends Go to the Beach	soaked, gorgeous	decide
Marquez Monkeys Around	reckless, ruin	quickly
If Elephants Could Fly	imagine, soar	gaze
Leo Loses His Roar	ill, comfort	practice
Ellie Gets Stuck	leap, pause	notice
A New Jungle Friend	speedy, wise	carefully
Marquez's Backwards Day	ridiculous, tumble	difficult

Note: Embedded lessons were provided for targeted vocabulary words and participants were assessed on all targeted vocabulary words. Untaught words were included in the text of all nine stories but no embedded lessons were provided.

in early childhood classroom settings and shows promise as a low-cost preschool language intervention.

Acknowledgements

This research was supported by Cooperative Agreement R324C080011 (Center for Response to Intervention in Early Childhood) from the U.S. Department of Education, Institute of Education Sciences. We would like to thank the teachers and children who made this study possible.

Appendix A.

Table A1.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecresq.2014.12.004.

References

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal*, *107*(3), 251–271. http://dx.doi.org/10.1086/511706

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford.

Berkeley, S., Bender, W. N., Gregg Peaster, L., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities*, *42*(1), 85–95. http://dx.doi.org/10.1177/0022219408326214

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology*, *98*, 44–62.

Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research*, *72*(1), 101–130. http://dx.doi.org/10.3102/00346543072001101

Boehm, A. (2000). *Boehm test of basic concepts* (3rd ed.). San Antonio, TX: Pearson.

Bradfield, T. A., Besner, A. C., Wackerle-Hollman, A. K., Albano, A. D., Rodriguez, M. C., & McConnell, S. R. (2013). Redefining individual growth and development indicators: Oral language. *Assessment for Effective Intervention*, *39*(4), 233–244. http://dx.doi.org/10.1177/1534508413496837

Brett, A., Rothlein, L., & Hurley, M. (1996). Vocabulary acquisition from listening to stories and explanations of target words. *Elementary School Journal*, 415–422, doi: 0013-5984/96/9604-0003$01.00.

Center for Response to Intervention in Early Childhood (CRTIEC). (2009). Work scope of the Center for Response to Intervention in Early Childhood (CRTIEC). In M. R. Coleman, F. P. Roth, & T. West (Eds.), *Roadmap to pre-K RTI* (pp. 13–14). New York, NY: National Center for Learning Disabilities. Retrieved from http://www.ncld.org

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, *21*(8), 1106–1116. http://dx.doi.org/10.1177/0956797610375449

Coyne, M., McCoach, D., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly*, *30*, 74–88. http://dx.doi.org/10.2307/30035543

Coyne, M., McCoach, D., Loftus, S., Zipoli, R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *Elementary School Journal*, *110*, 1–18. http://dx.doi.org/10.1086/598840

Coyne, M., Simmons, D., Kame'enui, E., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality*, *12*, 145–162. http://dx.doi.org/10.1207/s15327035ex1203_3

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Jr., Ruby, M., Crevecoeur, Y. C., et al. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness*, *3*(2), 93–120. http://dx.doi.org/10.1080/19345741003592410

Desmarais, C., Nadeau, L., Trudeau, N., Filiatrault-Veilleux, P., & Maxès-Fournier, C. (2013). Intervention for improving comprehension in 4–6 year old children with specific language impairment: Practicing inferencing is a good thing. *Clinical Linguistics & Phonetics*, *27*(6–7), 540–552. http://dx.doi.org/10.3109/02699206.2013.791880

Diamond, K., & Powell, D. (2011). An interactive approach to the development of a professional development intervention for Head Start teachers. *Journal of Early Intervention*, *33*, 75–93. http://dx.doi.org/10.1177/1053815111400416

Dickinson, D. (2011). Teachers' language practices and academic outcomes of preschool children. *Science*, *333*(6045), 964–967. http://dx.doi.org/10.1126/science.1204526

Dickinson, D., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly*, *22*, 243–260. http://dx.doi.org/10.1016/j.ecresq.2007.03.001

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

Dunn, L. M., & Dunn, L. (1996). *Peabody picture vocbulary test-III* (3rd ed.). Circle Pines, MN: American Guidance Service.

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1–44. http://dx.doi.org/10.1080/19345740802539200

Elley, W. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*, 174–187. http://dx.doi.org/10.2307/747863

Fischel, J. E., & Landry, S. H. (2008). *Impact of language-enhancement interventions on young children's early literacy skills. Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.

Foorman, B., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice*, *16*, 203–212. http://dx.doi.org/10.1111/0938-8982.00020

Fuchs, D., & Fuchs, L. (2006). New directions in research, introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*, 93–99.

Gersten, R., Compton, D., Connor, C., Dimino, J., Santoro, L. E., Linan-Thompson, S., et al. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in primary grades. A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Science, U.S. Department of Education.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*(2), 149–165.

Greene-Brabham, E., & Lynch-Brown, C. (2002). Effects of teachers' reading-aloud styles on vocabulary acquisition and comprehension of students in the early elementary grades. *Journal of Educational Psychology*, *94*(3), 465–473. http://dx.doi.org/10.1037//0022-0663.94.3.465

Greenwood, C., Bradfield, T., Kaminski, R., Linas, M., Carta, J., & Nylander, D. (2011). The response to intervention approach in early childhood. *Focus on Exceptional Children*, *43*(9), 1–22.

Greenwood, C., Carta, J., Atwater, J., Goldstein, H., Kaminski, R., & McConnell, S. (2012). Is a response to intervention (RTI) approach to preschool language and early literacy instruction needed? *Topics in Early Childhood Special Education*, http://dx.doi.org/10.1177/0271121412455438

Greenwood, C., Carta, J., Spencer, E., Guerrero, G., Kong, N., Atwater, J., et al. (in press). The effects of a Tier 2 vocabulary and comprehension storybook intervention on preschool children's early learning: A replication. *The Elementary School Journal*.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*, 165–180. http://dx.doi.org/10.1177/001440290507100203

Justice, L., Mashburn, A., Hamre, B., & Pianta, R. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, *23*(1), 51–68. http://dx.doi.org/10.1016/j.ecresq.2007.09.004

Justice, L., Mashburn, A., Pence, K. L., & Wiggins, A. (2008). Experimental evaluation of a preschool language curriculum: Influence on children's expressive language skills. *Journal of Speech Language, and Hearing Research*, *51*, 983. http://dx.doi.org/10.1044/1092-4388(2008/072)

Justice, L., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools*, *36*, 17–32. http://dx.doi.org/10.1044/0161-1461(2005/003)

Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, *31*(3), 259–272. http://dx.doi.org/10.1111/j.1467-9817.2008.00370.x

Kennedy, C. (2005). *Single case designs for educational research*. Boston, MA: Pearson.

Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening

comprehension in 4–6-year-old children. *Reading Research Quarterly*, *47*(3), 259–282. http://dx.doi.org/10.1002/rrq.020

Loftus, S., Coyne, M., McCoach, B., & Zipoli, R. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research & Practice*, *25*(3), 124–136. http://dx.doi.org/10.1111/j. 1540-5826.2010.00310.x

Lonigan, C., Phillips, K., Cantor, B., Anthony, J., & Goldstein, H. (2003). A computer assisted instruction phonological sensitivity program for preschool children at risk for reading problems. *Journal of Early Intervention*, *25*, 248–262. http://dx.doi.org/10.1177/105381510302500402

Lonigan, C., Schatschneider, C., & Westburg, L. (2008). *Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling. Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.

Lonigan, C., Shanahan, T., & Cunningham, A. E. (2009). Impact of shared-reading interventions on young children's early literacy skills. In *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *80*(3), 300–335. http://dx.doi.org/10.3102/0034654310377087

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York, NY: Cambridge University Press.

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*(2), 979–1007. http://dx.doi.org/10.3102/0034654309332561

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development*, *19*, 7–26. http://dx.doi.org/10.1080/10409280701838603

Neuman, S. B., & Kaefer, T. (2013). Enhancing the intensity of vocabulary instruction for preschoolers at risk. *Elementary School Journal*, *113*(4), 589–608. http://dx.doi.org/10.1086/669937

Pence, K. L., Justice, L., & Wiggins, A. K. (2008). Preschool teachers' fidelity in implementing a comprehensive language-rich curriculum. *Language, Speech, and Hearing Services in Schools*, *39*, 329–341. http://dx.doi.org/10.1044/0161-1461(2008/031)

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology*, *94*(1), 23–33. http://dx.doi.org/10.1037//0022-0663.94.1.23

Penuel, W. R., Bates, L., Gallagher, L. P., Pasnik, S., Llorente, C., Townsend, E., et al. (2012). Supplementing literacy instruction with a media-rich intervention: Results of a randomized controlled trial. *Early Childhood Research Quarterly*, *27*(1), 115–127. http://dx.doi.org/10.1016/j.ecresq.2011.07.002

Piasta, S. B., Justice, L. M., Cabell, S. Q., Wiggins, A. K., Turnbull, K. P., & Curenton, S. M. (2012). Impact of professional development on preschool teachers' conversational responsivity and children's linguistic productivity and complexity. *Early Childhood Research Quarterly*, *27*(3), 387–400. http://dx.doi.org/10.1016/j.ecresq.2012.01.001

Pollard-Durodola, S. D., Gonzalez, J. E., Simmons, D. C., Kwok, O., Taylor, A. B., Davis, M. J., et al. (2011). The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. *Exceptional Children*, *77*(2), 161–183. http://dx.doi.org/10.1177/001440291107700202

Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. D. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research & Practice*, *25*(3), 110–123. http://dx.doi.org/10.1111/j. 1540-5826.2010.00309.x

Qi, C., Kaiser, A., Milan, S., & Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the PPVT-III. *Language, Speech, and Hearing Services in Schools*, *37*, 5–16. http://dx.doi.org/10.1044/0161-1461(2006/002)

Robbins, C., & Ehri, L. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology*, *86*, 54–64. http://dx.doi.org/10.1037/0022-0663.86.1.54

Segers, E., & Verhoeven, L. (2004). Computer-supported phonological awareness intervention for kindergarten children with specific language impairment. *Language, Speech, and Hearing Services in Schools*, *35*, 229–239. http://dx.doi.org/10.1044/0161-1461(2004/022)

Segers, E., & Verhoeven, L. (2003). Effects of a vocabulary training by computer in kindergarten. *Journal of Computer Assisted Learning*, *19*, 557–566. http://dx.doi.org/10.1046/j.0266-4909.2003.00058.x

Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language*, *24*, 123–138.

Sénéchal, M., Thomas, E., & Monker, J. (1995). Individual differences in 4-year-olds' ability to learn new vocabulary. *Journal of Educational Psychology*, *87*, 218–229. http://dx.doi.org/10.1037/0022-0663.87.2.218

Shamir, A., Korat, O., & Fellah, R. (2010). Promoting vocabulary, phonological awareness and concept about print among children at risk for learning disability: Can e-books help? *Reading and Writing*, *25*(1), 45–69. http://dx.doi.org/10.1007/s11145-010-9247-x

Spencer, E., Goldstein, H., & Kaminski, R. (2012). Teaching vocabulary in storybooks: Embedding explicit vocabulary instruction for young children. *Young Exceptional Children*, *15*, 18–32. http://dx.doi.org/10.1177/1096250611435367

Spencer, E., Goldstein, H., Sherman, A., Noe, S., Tabbah, R., Ziolkowski, R., et al. (2013). Effects of an automated vocabulary and comprehension intervention: An early efficacy study. *Journal of Early Intervention*, *34*, 195–221. http://dx.doi.org/10.1177/1053815112471990

Spencer, T., Goldstein, H., Kelley, E. S., Sherman, A., & McCune, L. (2015). *A curriculum-based measure of language comprehension for preschoolers: Reliability and validity of the Assessment of Story Comprehension (ASC)*. Manuscript submitted for publication.

Spencer, T. D., & Goldstein, H. (2011, July). The assessment of story comprehension (ASC): A preliminary investigation of reliability and validity. In S. Piasta (Ed.), *New measures for investigation emergent literacy environments and skill development. Symposium presented at the annual meeting of the Society for the Scientific Study of Reading*.

Tompkins, V., Zucker, T. A., Justice, L. M., & Binici, S. (2013). Inferential talk during teacher–child interactions in small-group play. *Early Childhood Research Quarterly*, *28*(2), 424–436. http://dx.doi.org/10.1016/j.ecresq.2012.11.001

Ukrainetz, T. (2006). The implications of RTI and EBP for SLPs: Commentary on L.M. Justice. *Language Speech and Hearing Services in Schools*, *37*, 298–303. http://dx.doi.org/10.1044/0161-1461(2006/034)

van Daal, V., & Sandvik, J. (2013). The effects of multimedia on early literacy development of children at risk: A meta-analysis. In A. Shamir, & O. Korat (Eds.), *Technology as a Support for Literacy Achievements for Children at Risk* (pp. 73–119). Dordrecht: Springer.

van Kleeck, A. (2006). Fostering inferential language during book sharing with pre-readers: A foundation for later text comprehension strategies. In A. van Kleeck (Ed.), *Sharing books and stories to promote language and literacy* (pp. 269–317). San Diego, CA: Plural Publishing.

van Kleeck, A., Gillam, R., Hamilton, L., & Cassandra, M. (1997). The relationship between middle-class parents' book-sharing discussion and their preschoolers' abstract language development. *Journal of Speech, Language, and Hearing Research*, *40*, 1261–1271. http://dx.doi.org/10.1044/jslhr.4006.1261

van Kleeck, A., van der Woude, J., & Hammett, L. (2006). Fostering literal and inferential skills in Head Start preschoolers with language impairments using scripted book-sharing discussions. *American Journal of Speech-Language Pathology*, *15*, 85–94. http://dx.doi.org/10.1044/1058-0360(2006/009)

VanDerHeyden, A. M., Snyder, P. A., Broussard, C., & Ramsdell, K. (2008). Measuring response to early literacy intervention with preschoolers at risk. *Topics in Early Childhood Special Education*, *27*(4), 232–249. http://dx.doi.org/10.1177/0271121407311240

Verhallen, M., & Bus, A. (2010). Low-income immigrant pupils learning vocabulary through digital picture storybooks. *Journal of Educational Psychology*, *102*(1), 54–61. http://dx.doi.org/10.1037/a0017133

Verhallen, M., Bus, A., & de Jong, M. (2006). The promise of multimedia stories for kindergarten children at risk. *Journal of Educational Psychology*, *98*(2), 410–419. http://dx.doi.org/10.1037/0022-0663.98.2.410

Washington, J., & Craig, H. (1999). Performances of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test-III. *Language, Speech, and Hearing Services in Schools*, *30*, 75–82. http://dx.doi.org/10.1044/0161-1461.3001.75

Wasik, B., & Bond, M. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology*, *93*, 243–250. http://dx.doi.org/10.1037/0022-0663.93.2.243

Wasik, B. A., & Hindman, A. H. (2014). Understanding the active ingredients in an effective preschool vocabulary intervention: An exploratory study of teacher and child talk during book reading. *Early Education and Development*, 1–22. http://dx.doi.org/10.1080/10409289.2014.896064 (Ahead of print)

Whitehurst, G., Arnold, D., Epstein, J., Angell, A., Smith, M., & Fischell, J. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology*, *30*, 679–689. http://dx.doi.org/10.1037/0012-1649.30.5.679

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical evaluation of language fundamentals preschool* (2nd ed.). San Antonio, TX: Harcourt Assessments.

Wright, T. S. (2012). What classroom observations reveal about oral vocabulary instruction in kindergarten. *Reading Research Quarterly*, *47*(4), 353–355. http://dx.doi.org/10.1002/RRQ.026

Zucker, T. A., Justice, L., Piasta, S. B., & Kaderavek, J. N. (2010). Preschool teachers' literal and inferential questions and children's responses during whole-class shared reading. *Early Childhood Research Quarterly*, *25*(1), 65–83. http://dx.doi.org/10.1016/j.ecresq.2009.07.001

Zucker, T. A., Solari, E. J., Landry, S. H., & Swank, P. R. (2013). Effects of a brief tiered language intervention for prekindergartners at risk. *Early Education & Development*, *24*(3), 366–392. http://dx.doi.org/10.1080/10409289.2012.664763