Communication Sciences and Disorders Faculty Publications

Communication Sciences and Disorders

6-2016

Systematic Replication of the Effects of a Supplementary, Technology-assisted, Storybook Intervention for Preschool Children with Weak Vocabulary and Comprehension Skills

Charles R. Greenwood
University of Kansas

Judith J. Carta
University of Kansas

Elizabeth S. Kelley
University of Missouri-Columbia

Gabriela Guerrero
University of Kansas

Na Young Kong
University of Missouri-St. Louis

See next page for additional authors

Follow this and additional works at: http://scholarcommons.usf.edu/csd_facpub

Part of the Communication Sciences and Disorders Commons

Authors

Charles R. Greenwood, Judith J. Carta, Elizabeth S. Kelley, Gabriela Guerrero, Na Young Kong, Jane Atwater, and Howard Goldstein

SYSTEMATIC REPLICATION OF THE EFFECTS OF A SUPPLEMENTARY, TECHNOLOGY-ASSISTED, STORYBOOK INTERVENTION FOR PRESCHOOL CHILDREN WITH WEAK VOCABULARY AND COMPREHENSION SKILLS

ABSTRACT

In 2013, Spencer, Goldstein, Sherman, et al. reported the promising effects of a supplemental, technology-assisted, storybook intervention (Tier 2) containing embedded instruction targeting the oral language learning of preschool children at risk for delays. We sought to advance knowledge of the intervention by replicating it in a new sample and examining children's responses to the narrator's instructional prompts and associations with learning outcomes. Results indicated that children were highly successful in responding with the narrator's task-management prompts (i.e., *turn the page*), particularly after the first book. Children were much less proficient in correctly responding to the narrator's word-teaching prompts (i.e., "*say enormous*"), but improved over additional storybooks. Exposure to the intervention accelerated children's weekly oral language learning, and effect sizes were comparable to those of Spencer et al. Children's increased word knowledge was positively correlated with their correct responding to the narrator's word-teaching prompts in particular. Implications for research and practice are discussed.

Charles R. Greenwood
Judith J. Carta
UNIVERSITY OF KANSAS

Elizabeth S. Kelley
UNIVERSITY OF
MISSOURI–COLUMBIA

Gabriela Guerrero
UNIVERSITY OF KANSAS

Na Young Kong
UNIVERSITY OF
MISSOURI–ST LOUIS

Jane Atwater
UNIVERSITY OF KANSAS

Howard Goldstein
UNIVERSITY OF
SOUTH FLORIDA

THIS manuscript reports results from a single-case experimental study of the effects of a promising supplemental language and early literacy storybook intervention for preschool children screened for weak skills and likely to benefit from additional instructional experiences. Impetus for the work was the Center for Response to Intervention in Early Childhood (CRTIEC), a research and development center supported by the National Center for Special Education Research, Institute for Education Sciences (IES), tasked with developing and validating supplemental (Tier 2) and individualized (Tier 3) interventions for children with weak and very weak language and literacy skills for use in a Multiple Tier System of Supports (MTSS) approach to services (Greenwood et al., 2014). Children with weak skills in this work were those scoring between 1 and 2 standard deviations below the normative mean. Those with very weak skills scored at or below 2 standard deviations. Children with Individual Education Programs (IEPs) were included in both Tier 2 and 3 based on their screening results.

The conceptual framework guiding CRTIEC's development of new evidence-based practices for these purposes consisted of several stages that organized iterative development with empirical replication research (IES Goal 2 research) (Greenwood et al., 2014; Horner et al., 2005), leading eventually to randomized trials evaluating the efficacy of the promising new interventions (IES Goal 3 research) (National Center for Special Education Research [NCSER], 2014). The earliest stage in the development process was the new intervention design informed by knowledge of existing evidence-based practices, the need to deliver with fidelity within real preschool classrooms in small groups, and the need for efficiency where these supplemental experiences would likely be supervised by relatively unskilled paraprofessionals rather than a teacher.

Thus, the first study in this sequence of development investigations, Spencer et al. (2013), combined storybook reading with electronic media to achieve automated delivery in a small-group format in the classroom. Automated delivery in this case refers to instruction mediated and assisted by electronic technology, specifically, prerecorded audio of story text and embedded lessons. The authors reported that this intervention was feasible, implementable with fidelity, and promising because moderate outcome effects were achieved by low-performing children. The current study is a systematic replication of Spencer et al. (2013) in that the intervention effects were evaluated in a new sample of participants and the study addressed some new issues concerning the automated intervention (Kelley & Goldstein, 2014; Kennedy, 2005; Kratochwill et al., 2010).

A sizable literature indicates that children's engaged behaviors during instruction are key mediators/moderators of the educational experience leading to the learning of new skills (Greenwood & Kim, 2012; Powell, Burchinal, File, & Kontos, 2008; Williford, Vick Whittaker, Vitiello, & Downer, 2013). We also know that instruction that provides frequent opportunities for children to respond and that also models correct responding is predictive of (Hindman, Connor, Jewkes, & Morrison, 2008; Williford et al., 2013), or leads to, growth (Walsh & Blewitt, 2006) in children's vocabulary development. Spencer et al. (2013), however, did not examine the issue of engaging children's responding to opportunities provided by the prerecorded, narrated script during storybook reading lessons. Addressing this issue in the current study was deemed particularly important because there was no teacher present

for redirecting responding in the face of individual errors or the absence of responding. An understanding of the engagement behaviors of children during the intervention was expected to be helpful in improving the intervention and in planning future randomized efficacy trials. If variations in children's engagement are predictive of learning, such variations would be considered a moderator of children's outcomes. Alternatively, engagement may be considered a mediator of intervention effects because it is an active ingredient of the intervention in the production of outcomes. Also, anticipating future efficacy trials, the engagement and outcome data would need to be examined at the small-group (within-classroom level) and not only at the child level because the appropriate intervention unit of randomization in a controlled trial would most likely be the small-group/classroom level. The primary goal of the study was examination of child engagement and its relationship to child outcomes in small groups within classrooms.

Why Language and Literacy?

Important foundations of language and literacy are established in the preschool years. In particular, oral language skills in early childhood are well-documented predictors of later reading ability (Shanahan & Lonigan, 2008). Oral language skills are recommended target outcomes for most national and state early literacy frameworks in Head Start (Head Start, 2010), IDEA Part B–preschool (ECO Center, 2011), and state prekindergarten programs (Pre-K) (Barnett, Epstein, Friedman, Sansanelli, & Hustedt, 2011). Vocabulary knowledge enables children to label and understand the world, communicate with others, and achieve goals (Biemiller, 2006). Comprehension enables young children to act on what is said to them, and to learn and use new concepts as well as develop a foundation for learning to understand written text in the future (e.g., van Kleeck, 2008). Unfortunately, having limited oral language skills places children at high risk for later reading failure (Catts, Fey, Zhang, & Tomblin, 2001). Thus, effective oral language and comprehension instructional interventions prior to kindergarten are critically important to improving results for many children.

Substantial efforts have been devoted to improving the quality of oral language and literacy instruction in early childhood settings with some promising results. Positive effects are associated with intensive professional development programs and high-fidelity implementation of curricula in early childhood classrooms (Tier 1) (Dickinson & Caswell, 2007; Justice, Mashburn, Pence, & Wiggins, 2008; Wasik, Bond, & Hindman, 2006). Unfortunately, these are not often used in typical programs, and what is used may not be sufficient to close achievement gaps for students at risk (Greenwood & Kim, 2012; Justice, Hamre, & Pianta, 2008). For children with weak or limited language skills, a more intensive approach may be necessary.

Can Effectiveness be Increased Using Multitiered Systems of Support?

The preschool sector, like K–5 schooling, appears increasingly interested in using prevention-oriented approaches to intervene earlier with specific children in order

to address delays and prevent later learning disabilities (Buysee & Peisner-Feinberg, 2013; Johnston, 2011). Multitiered System of Support (MTSS) or Response to Intervention (RTI) are early intervening approaches that (*a*) identify students not making progress in the core curriculum (Tier 1), and (*b*) provide them with access to additional supplementary (Tier 2) or individualized instruction (Tier 3) to accelerate their progress (Johnston, 2011). Typically in MTSS, seasonal (i.e., fall, winter, spring) universal screening focused on desired instructional outcomes is used to identify children who are not making progress and are likely to benefit from more intensive instruction. Also in MTSS, special education services are not necessarily bundled with Tier 2 or Tier 3 instruction, depending on the outcomes of interest. Children with mobility issues may well be supported for literacy learning in Tier 1, 2, or 3, irrespective of their Individual Education Program (IEP) goals. Inversely, children with language and literacy delays and an IEP with literacy goals may be receiving Tier 2, and not Tier 3, because of the child's level of progress.

Making literacy instruction more intensive in Tiers 2 and 3 is often achieved by manipulating multiple factors, and there is no universal agreement as yet on how best to increase intensity. Typical factors range from decreasing the teacher-pupil ratio by using small groups (Tier 2) or one-on-one instruction (Tier 3) and at the same time adding more time per day for students to learn the subject matter. Instructional factors also are used in combination with changes in grouping and increased time. These include making instruction in Tiers 2 and 3 more explicit, rather than implicit, as is more common in Tier 1. Increasing opportunities to respond, correction and feedback, and scripting teacher-student interactions around content tasks are all possible factors, among others (e.g., general case design, scaffolding, etc.).

Are Supplemental, Intensive (Tier 2) Preschool Language Interventions Efficacious?

One particular barrier to preschool MTSS is the lack of evidence-based language and literacy instruction for at-risk preschoolers with weak (Tier 2) and very weak (Tier 3) skills (Greenwood et al., 2011). Several research groups have examined supplemental language and literacy interventions, often as part of an MTSS approach. Gettinger and Stoiber (2007) provided supplemental instruction to preschool children with weak (Tier 2) and very weak skills (Tier 3) in the context of an MTSS model. Tier 1 curriculum was experienced by all children, supplemental Tier 2 instruction was taught in teacher-directed, small groups, and Tier 3 instruction consisted of individual tutoring provided by graduate students and volunteers. Similar findings have been reported by others (Landry, Assel, Anthony, & Swank, 2013; Pullen, Tuckwiller, Konold, Maynard, & Coyne, 2010; VanDerHeyden, Witt, & Gilbertson, 2007).

It appears that for young children with weak skills, the most effective supplementary instruction is likely to be intensive and explicit. In the vocabulary domain, children learn more words, and more about those words, when teaching is explicit than in conditions of incidental exposure (Coyne, McCoach, & Kapp, 2007; Pullen & Justice, 2003; Pullen et al., 2010). Intensive instruction, provided by repeated

readings or days of instruction, also provides an advantage for vocabulary learning relative to brief or single exposures (Beck & McKeown, 2007; Penno, Wilkinson, & Moore, 2002). Even in optimal conditions of intensive, explicit interventions, children generally learn only a portion of all words taught. Biemiller and Boote (2006) reported children learning 22% or 41% of word meanings in conditions of repeated readings and explicit teaching. Beck and McKeown (2007) reported learning of 30% of meanings taught. These findings illustrate the challenge of effective vocabulary instruction, when even well-designed interventions produce only moderate results.

Are Tier 2 Interventions Feasible to Implement in Real Preschool Classrooms?

In addition to efficacy, MTSS interventions need to be feasible for use in preschool classrooms, often within contexts with limited resources in terms of personnel, time, and professional preparation. Also, a substantial body of research has highlighted the challenge of changing teacher practices with respect to language and literacy instruction. Justice, Hamre, and Pianta (2008) reported that while 92% of teachers completed all lesson components, the same teachers struggled to provide high-quality language modeling—just 8% of observed sessions. Dickinson (2011) reported similarly discouraging results. Teachers included just 38% of instructional elements during book reading. To provide the intensive, explicit instruction critical to the success of students with weak and very weak skills, there is a need to ensure that effective interventions can be implemented with high fidelity and dosage.

The use of technology is an intriguing option for intervention delivery in early childhood classrooms. Reviews of the effects of computer-assisted approaches have indicated an overall positive but small effect on the learning of early language and literacy skills (Blok, Oostdam, Otter, & Overmaat, 2002; van Daal & Sandvik, 2013). Automated approaches have been effective in improving reading fluency and reading comprehension (Gibson, Cartledge, & Keyes, 2011; Savage, Abrami, Hipps, & Deault, 2009), phonological awareness (Gillam et al., 2008; Lonigan, Phillips, Cantor, Anthony, & Goldstein, 2003; Segers & Verhoeven, 2004; Wild, 2009), and phonics skills (Macaruso & Walker, 2008). Research groups have specifically examined the benefits of automated programs for children at risk of later reading disability (Chambers, Slavin, & Madden, 2011; Lonigan et al., 2003), and some have reported that children with the poorest preintervention skills have benefited the most (Macaruso & Walker, 2008). There is also evidence that computer-assisted instruction can be effectively administered in small groups—an important consideration for feasibility of implementation (Chambers et al., 2011).

Another technology contribution is the potential for high-fidelity implementation. For example, an automated program can provide scripted, explicit instruction, consistent instructional language, and consistent dosage. Automated programs can be designed to include critical elements of effective instruction, such as multiple opportunities to respond for each child, which may be more difficult to achieve during small-group instruction delivered by a teacher. Automated inter-

ventions have the potential to effectively and efficiently supplement high-quality instruction at Tier 1, as well as provide intensive, explicit instruction for a select group of students who may benefit from additional instruction. Assisted delivery through electronic technology places little demand on staff resources (e.g., does not require extensive training, resources, or expertise) and has the potential to overcome these challenges to implementing supplementary instruction in preschool programs.

Spencer et al. (2013) took this approach by producing a series of instructional storybooks containing embedded, explicit instruction delivery by an assisted technology system for use in a Tier 2 supplemental MTSS intervention approach. As expected, high fidelity and dosage of the intervention were observed as implemented in the absence of direct teacher participation. Also reported were systematic improvements in children's vocabulary and comprehension in a single-case experiment with nine students identified as having weak skills (Spencer, Goldstein, & Kaminski, 2012), and that only words that received explicit instruction in the scripted lessons, not those just heard in the story (i.e., control words), were learned.

The aim of this research was to systematically replicate the technology-assisted storybook intervention of Spencer et al. (2013) with an additional focus on the role of child engagement and its association to learning outcomes. Our hypothesis was that task management would be important to learning, but that responding correctly to the narrator's word-teaching prompts would be even more important. A secondary aim was to strengthen the external validity of this promising intervention through replication by others (Horner et al., 2005). Our questions were: (1) How accurately did children respond to the narrated instructional audio, and were there accuracy differences in task management (e.g., *turn the page*) versus word-teaching prompts (e.g., *say ridiculous*)? (2) What were the effects of the intervention package on children's acquisition of new vocabulary and proficiency in answering questions about the stories? (3) Was children's accuracy in responding to the narrator's word-teaching instruction differentially associated with their vocabulary and story comprehension learning outcomes?

Method

Participants

A sample of three preschool teachers/classrooms and nine children (three per classroom) served in an early learning center in an urban district in eastern Kansas were enrolled in this study. Four- and 5-year-old children in their year before kindergarten were recruited for participation, and parental consent was sought. Children's classroom programs were half-day, 4 days per week. Children ranged in age from 53 to 62 months ($M = 57.8$, $SD = 3.3$). One child was White, five Hispanic, one Asian-American, and two African-American. Teachers reported that three of the nine children had an IEP and three of the nine were English language learners (ELL). No children had IEP and were also ELL. ELL children on the CELF-P2 language measure ($M = 70.3$) and those with IEPs ($M = 63.3$) performed more poorly than did the students without these characteristics ($M = 84.0$). This language-skill difference was also seen on the PPVT with ELL children, who were

lower, at 84.3, than either children who were English only and receiving no special education services (87.7) or children with IEPs (88.7).

All consented children were universally screened for weak oral language skills using the process employed by Spencer et al. (2013). Measures included (*a*) an experimental universal screener for expressive vocabulary (Picture Naming, IGDI 2.0) (Bradfield et al., 2013) and (*b*) two standardized language tests for additional information, the Peabody Picture Vocabulary Test (PPVT-IV) (Dunn & Dunn, 2007) and the Comprehensive Evaluation of Language Fundamentals (CELF-P2) (Wiig, Secord, & Semel, 2004).

A cut point of 10 out of 15 correct responses, a level at 70% classification accuracy on the Picture Naming measure, was used to identify children with skills below average (Bradfield et al., 2013). The mean Picture Naming screening score of children was 7.3 ($SD = 2.5$). As a group, these children were nearly a standard deviation below the mean in receptive vocabulary (PPVT-IV) and nearly 2 standard deviations below the mean in total language proficiency (CELF-P2) (see Table 1). Only Child 5 had a screening score above the 10 point cut, but was included because of a low CELF-P2 score.

Design

The repeated-acquisition, single-case design (RAD) was used (Kennedy, 2005, p. 163; Spencer et al., 2013) because the research questions considered the issue of repeated acquisition of new content in a series of storybooks delivered to small groups in classrooms under different conditions. The RAD is particularly well suited to addressing questions of academic skills acquisition as well as other educationally relevant skills as a function of alternative instructional interventions (Riley-Tillman & Burns, 2009). Acquisition of new word skills assessed before

Table 1. Participant Characteristics

Classroom and Child	Age (Mos.)	Gender	ELL	IEP	Picture Naming 2.0	PPVT-IV	CELF-P2
Classroom 1:							
Child 1	54	M	Yes	No	5	78	69
Child 2	62	F	Yes	No	7	92	69
Child 3	56	M	No	No	9	107	102
Classroom 2:							
Child 4	59	M	No	Yes	5	75	50
Child 5	62	M	No	Yes	13	92	65
Child 6	60	M	No	No	6	83	83
Classroom 3:							
Child 7	53	M	No	Yes	8	99	75
Child 8	58	M	Yes	No	6	83	73
Child 9	56	F	No	No	7	73	67
M	57.8				7.3	86.9	72.6
SD	3.3				2.5	11.4	14.2
Min	53				5	73	50
Max	62				13	107	102

Note.—There were nine children: 7 male, 2 female; 3 ELL, 3 IEP. Age is reported in months and is the child's age at the beginning of the study. ELL = English language learner, IEP = Individualized Education Plan; Picture Naming IGDI 2.0 had a maximum of 15 and a fall cut-off score of 10. PPVT-IV = Peabody Picture Vocabulary Test, 4th ed. (Dunn & Dunn, 2007); CELF-P2 = Clinical Evaluation of Language Fundamentals Preschool, 2nd ed. (Wiig, Secord, & Semel, 2004).

and after instruction in each storybook in this study was used in the RAD to produce repeated evidence that the treatment was responsible for new acquisition within each storybook. Consequently, reversals or withdrawal of the intervention conditions, typical of many single-case designs, are not required or expected in the RAD (Kennedy, 2005). Where confidence in the repeatability of the treatment effect is demonstrated by at least three replications of the treatment effect within participants in most single-case designs (Kratochwill et al., 2010), the RAD design provided nine such opportunities in this investigation, one for each book in the series.

As do most single-case designs, RAD controls for most of the expected threats to internal validity. For example, maturation and history are controlled by the repeated demonstrations of effect in the presence of the intervention condition. The primary threat to internal validity in the RAD is related to the equivalence of the different content tasks used as the dependent variables in the design (Kennedy, 2005). Another is potential sequence or order effects. Evidence of an equivalence problem in this study would be indicated by particularly low or high performance in one or several storybooks compared to the others. A sequence problem would be suggested if an increasing trend in performance across the series of storybooks was observed, suggesting that learning in earlier books was facilitating learning in the later books. As regards controlling word equivalence in the design, a pool of words was created from a compendium of adult vocabulary (see below). Words were selected for their educational value for the children but also because it was unlikely they would know them, thus providing a good preinstruction floor for demonstrating acquisition. Regarding sequence, the words were assigned to storybooks in no particular order and the books placed in quasi-random sequence.

Conditions in the RAD were A and B, where A = performance on the words to be taught and story questions *before* instruction (pretest assessment) and B = performance on the words taught *after* three instructional listening sessions per book (postinstruction assessment). Fidelity of implementation was assessed during all sessions. Effects were examined at both the child and classroom small-group levels, compared to only the child level in Spencer et al. (2013).

Setting

Assessment and intervention took place in the classrooms. Automated storybook listening stations for small-group instruction were set up in low-distraction classroom areas. However, competing noise and activity in these busy classrooms was the norm. Four research staff members, three graduate research assistants, and one coordinator facilitated delivery of the intervention, and three also served as data collectors. Due to time restrictions, it was not possible to keep these roles separate. Steps to reduce sources of bias are described below.

The Technology-Assisted, Storybook Intervention for Small Groups

We replicated the technology-assisted, storybook vocabulary and comprehension intervention of Spencer et al. (2013) for small groups for children with weak literacy skills (Tier 2) previously described. The original developers provided

the replication team with the intervention materials and the mastery assessments for each book used in the original study, and helped plan and train the implementation team using a mix of site visits and teleconferencing. The replication team met weekly to update children's progress graphs and work through procedural questions and issues. Some questions were asked and resolved by the original team.

The intervention materials included a series of 10 storybooks, the *Jungle Friends* (one initial training book and nine books containing embedded instruction), accompanying a digital audio narration, headphones, and MP3 player. Participants completed intervention in small groups of three children and an adult facilitator (one of the research staff, a research assistant). Children participated in three instructional sessions per week per storybook and were typically assessed on Friday. Exceptions were holidays or child absences. Missed sessions were made up.

The project coordinator taught the research assistants to perform the facilitator role and monitored fidelity of implementation. As part of training, the coordinator and facilitators reviewed and critiqued sample videos of listening centers. Facilitators and the implementation team held weekly meetings to discuss general recommendations for promoting task-appropriate engagement and how to deal with instances of challenging behavior if necessary.

A facilitator was present for each and every session; however, who the facilitator was varied session to session because of schedules. The facilitator's role was to set up, supervise, and clear up/put away the listening station when finished. Facilitators supervised children's participation in the session in several ways. They provided a brief review of the behavioral expectations of children in the session (e.g., *keep your headphones on*), and during the session reminded and redirected children when needed if they failed to turn to the correct page. Facilitators also used the headphones to make sure the audio story narration was being heard by all and that the children were having the opportunity to respond to the embedded lessons. Facilitators also responded to the narrator's prompts—for example, repeating a vocabulary word when prompted—just as the children did. However, facilitators did not provide any instruction, prompting, encouragement, or feedback outside of that provided by the narrator. Fidelity on facilitators' performance was assessed and monitored throughout the study by the project coordinator.

Instructional content design. The active ingredients in each storybook were the embedded, narrated lessons focused on teaching two new vocabulary words and improving children's ability to answer questions about the book. On the prerecorded audio, the narrator read the story text aloud and presented embedded instruction using systematic instructional language during brief interruptions to the story text. All children received the same experiences in the intervention, and the curriculum was not adapted for ELL students or children with IEPs.

Selection of vocabulary targets and design of instruction followed the approach for robust vocabulary instruction recommended by Beck and colleagues (Beck & McKeown, 2007). Words selected were those occurring somewhat frequently in adult conversation (e.g., *speedy, protect*) with potential to be useful in a variety of oral and written language contexts (e.g., *enormous*). Words were those that were unlikely to be already known by preschoolers, and were challenging words likely to require instruction in order to be learned. If words could be readily acquired through incidental exposure or brief instruction (e.g., *map*), they were excluded.

Only words that could be explained in the context of the story and associated with a key story event or character were included. The vocabulary words included 7 verbs and 11 adjectives. No nouns were included.

Embedded lessons for vocabulary provided explicit child-friendly definitions as well as opportunities to say the word and definition, and provided information about the word's meaning by using it in a number of contexts likely to be familiar to children (Spencer et al., 2012). Children were encouraged to respond, and a pause was provided as a prompt to evoke a response. For example, *Leo has to be brave. He doesn't want to be scared of the dentist. Brave. Say brave:* [pause for child response]. *Brave means not scared. What word means not scared?* [pause for child response] *Brave.* (For greater detail, see Spencer et al., 2013.) Two prerecorded lessons were provided for each word—a longer lesson embedded in the story text, and a briefer lesson immediately following the story.

Story-targeted comprehension instruction was taught, guided by the work of van Kleeck et al. (van Kleeck, 2006, 2008; van Kleeck, van der Woude, & Hammett, 2006). Comprehension questions targeted inferential reasoning (21 of 27), and six were recall. Inferential questions were questions for which the answers were not explicitly stated in the text. They were often *why* questions, and an appropriate response to the question required an inference using information in the story. For example, the question *Why was Ellie happy?* was inferential because there was no single statement in the text that explained that *Ellie was happy because . . . .* The embedded instruction in each story always included several possible reasons why Ellie might be happy (e.g., a first day of school, meeting new friends), and an appropriate response to this question could include one or more of these reasons. Other inferential questions targeted predictions about story content (e.g., *What do you think will happen in this story?*) or predictions about future events (e.g., *Do you think the Jungle Friends will go to the beach again?*). Recall questions were fact based from the story.

The embedded lesson for each story question occurred before the story; lessons for the other two questions occurred immediately after the story. Consistent instructional language also was provided for each comprehension lesson. First, the story question was introduced. Next, the narrator provided an appropriate answer and a "think aloud" explanation for the answer. For example: *At the end of the story, Ellie was happy. Why was Ellie happy?* [pause for child response] *Because she made some new friends. I bet she felt happy to have new friends.*

Measurement and Reliability

Intervention fidelity and dosage. A checklist was developed and administered by the facilitators to measure exposure to the intervention. Assessed were: (*a*) each child has a book, (*b*) each child has headphones, (*c*) the correct lesson is playing and audio levels are correctly set, (*d*) each child is reminded of the listening-center rules, and (*e*) each child is provided with reinforcement (e.g., stamps) for participating correctly. Fidelity was 100% for all listening sessions. Weekly in-person fidelity observations were conducted by the coordinator to assess each facilitator's fidelity with their prescribed role. All facilitators' evaluations were 100%. Intervention sessions lasted for 15 minutes. All missed sessions were made up to maintain

dosage at three listens per book. Because of a later start in the school year, children in Classroom 3 completed only seven of the nine books.

Student engagement with narrator's instruction. An observational checklist based on each book's recorded script was used for recording children's responding to the narrator's audio prompts for a response. Response to comprehension prompts was not assessed because of differences in testing and the assumption that engagement could be inferred from responsiveness to the word-teaching prompts. Prompts were classified as either task management (e.g., *Put your finger on . . .*) or word teaching (e.g., *What are things that leap?*). Task prompts were scored as completed or not completed, while word-teaching prompts were scored as either correct or incorrect/no response. The total number of prompts in the scripts and on the checklist for a lesson ranged from 20 to 29 (task-management) and 14 to 22 (word-teaching) prompts per book, respectively. Scores were a simple percentage calculated by dividing the number complete/correct by the total prompts for that book.

Observers were trained to complete these observations by the project coordinator. Training included reading a document containing the scoring rules and meeting with the project coordinator to discuss and answer questions. After training, observers and the project coordinator met weekly to provide opportunities to address and resolve any questions that emerged. Because staff resources were not sufficient for observing all students on each listening during the study, a sampling plan of the nine children and nine books was implemented with the goal of distributing them across all children books and listens within intervention books. The mean number of observations collected per child was four, ranging from one to six, for a total of 33.

Interobserver agreement scoring of student engagement. Interobserver agreement (IOA) was sampled from observations of listening session videotapes of four books (e.g., *Ellie Gets Stuck*, etc.). Videotape rather than in situ observation was used to reduce the intrusiveness of adding more adults to small classrooms and to save staff travel time. Observers recorded the accuracy of a single target child's response to all scripted task-management and target-content prompts. Percentage agreement was calculated by dividing the number of agreements by the total number of narrator prompts times 100. This calculation was made for the six unique pairings of observers (e.g., scores from observer 1 vs. 2, observer 1 vs. 3, etc.). The total number of narrators' prompts ranged from 26 to 28 for task management and from 19 to 25 for word teaching across the four books. The mean agreement overall was 91% ($SD = .8$) for task management and 86% ($SD = .7$) for word teaching. Similar values across the four books ranged from 80% to 97% for task management and from 76% to 91% for word teaching.

Vocabulary and comprehension mastery probes. Dependent variables were children's scores on nine mastery probes for vocabulary and comprehension learning. Mastery probes were designed to assess students' learning of the instructional targets taught in each book. Mastery probes were administered individually by a familiar examiner, one of the listening center facilitators, both before and after the third and final listen for each book. The listening center schedule was as follows: (1) children received a pretest on the book they were about to listen to, (2) they listened to the same book for three sessions per book, (3) they received

the posttest (identical to the pretest). Exceptions to the schedule reflected individual absences or weekly schedule changes. Make-up sessions were conducted to ensure that participants always listened to the book three times before receiving the posttest.

Items on the mastery probes were the same words that were taught during sessions. For each of the two vocabulary words taught per book, children were assessed on their improvement in definitional knowledge. Children were asked to define each target word (e.g., *What does "enormous" mean?*). These definitional items were scored using a three-point differential, reflecting complete knowledge (2 points = definition taught, other definition or synonym, definition plus an example), partial knowledge (1 point = fractional use in a phrase, meaningful use in a phrase or sentence, use in the story context or specific example), or no knowledge at all (0 points = incorrect or no response) (see Spencer et al., 2013). Point scores for definitional vocabulary items were summed for each book, and the total possible point score was equal to 2 words × 2 points or 4 points per book (see Spencer et al., 2013, p. 12). A score of 2 points was considered to be a clinically significant intervention effect because it was equal to either one word learned with full knowledge or two words learned with partial knowledge.

For story comprehension, mastery-monitoring probes were also administered before any listening to the story and again after three listens to the embedded instruction. Assessed were three comprehension questions tapping what was targeted in the intervention (e.g., *Why was Ellie happy?*). The comprehension probe questions were exact parts of the explicit instruction for each book, but only one of five parts of the explicit instruction. The probe questions represented a challenging task because a significant time lag occurred between their last listening and this probe question and because the question was now decontextualized from the storybook. Responses to each question were scored as (*a*) complete and accurate (2 points, when the child used information from the story or their own knowledge), (*b*) partially complete and accurate (1 point, when the child's response was general or vague, that is, it was related but unclear), or (*c*) incorrect (0 points, when the child was incorrect or no response was provided).

Like the vocabulary items, comprehension probe items were identical for both before and after occasions, with one exception. For five of the nine books, story questions tapped making predictions about story events. For these books, children were asked to respond to the prediction question after hearing the story once without embedded instruction (e.g., *What do you think will happen in this story?*). They were then assessed a second time after three listens to the story that included the embedded instruction. However, after instruction, the prediction question was pragmatically inappropriate because children already knew what would happen in the story. In these cases, children were asked to recall story events instead (e.g., *What happens in this story?*). Scores for comprehension items were summed for each occasion; total scores ranged from 0 to 6.

Interobserver agreement on scoring. Interobserver agreement (IOA) percentages were obtained for 75% of the mastery probes scored by two observers for the purpose of checking reliability. Exact score percentage agreement was equal to the [number of agreements / (number of agreements + number of disagreements) × 100]. IOA for vocabulary was 100% (SD = 0.0) for pretest and 99% (SD = 2.2) for

posttest. Similar values for the comprehension probes were 93% ($SD = 11.3$) for pretest and 91% ($SD = 12.1$) for posttest.

Statistical Data Analysis

A combination of graphical data display and descriptive statistics, including cross-classification with chi-square, Pearson r, and paired t-tests, was used to address the research questions. To address the first question, the child-engagement data were graphed across books overall, by classroom groups, and by individual children. This provided a look at replication of intervention effects by classrooms and children. To address the second question, child outcome data were analyzed descriptively and graphed in the style of the RAD to demonstrate repeated acquisition. The planned dataset consisted of nine children each with nine pretest (81) and nine posttest (81) scores, for a total of 162 child-test occasions. Incomplete data were explained by three children in Classroom 3 who heard only seven of the nine books, and one child in Classroom 2 who missed the Book 9 posttest. Thus, the total complete data analyzed were 149 child-test occasions.

The mean and standard deviation of pretest and posttest and gain in points were used overall, as well as by each consecutive book and by individual children. Chi-square was used to test the acquisition of new words learned. A new-words-learned variable was created based on children's points-earned scores, where one or more words learned was equal to 2 or more points (coded 1), indicating that they provided at least a complete definition of one word or a partial definition to two words, versus less than one word (coded 0). Chi-square for a 2 × 2 table was used to test the difference between two phases of the study (A = baseline [pretest] and B = treatment [posttest]) versus the words-learned outcome of 0 or 1.

We also used the Nonoverlap of All Pairs (NAP) effect-size index for single-case designs (Parker & Vannest, 2009). It was an appropriate choice because zero was a legitimate score for vocabulary and comprehension measures, and at pretest many children scored zero on each book. NAP was calculated by pooling the A ($n = 9$ books) and B ($n = 9$ books) phase point scores of all nine children in a Receiver Operating Curve (ROC) analysis (Parker & Vannest, 2009). To address engagement and outcomes relationships, Pearson r was used.

Results

Question 1: How Well Did the Children Respond to the Narrator's Task-Management versus Word-Teaching Prompts during Listening Sessions?

The mean percentage of compliance with the task-management prompts over all children and books was 88.5% ($SD = 16.4$). The same value for correct responding to the word-teaching prompts was lower, at 56.5% ($SD = 26.3$) (see Table 2). The trend in accurate task-management responding across books rose rapidly to high stable performance after the first book (varying from 82% to 98%, Books 2–9) (see upper left panel, Fig. 1). The corresponding trend in correct responding to word-teaching prompts started out low, at only 17.5% (Book 1), increasing linearly (slope = 8.4% gain per book), and reaching a high of 87.5% by Book 9.

Table 2. Percentage Task-Compliance and Word-Teaching Response Accuracy and Vocabulary and Comprehension Scores by Children (Sorted on Word Teaching)

						Vocabulary						Comprehension					
		%Task		%Word		Before		After		Gain		Before		After		Gain	
Student	Occasions	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1	5	58.2	14.0	35.4	14.1	.0	.0	.4	.9	.4	.9	.8	1.1	1.7	1.6	.9	1.1
4	4	93.2	6.0	48.0	35.4	.0	.0	.4	.9	.4	.9	1.1	1.7	1.6	1.0	.4	1.6
2	5	94.8	5.4	56.0	32.3	.0	.0	2.3	.9	2.3	.9	3.3	1.1	4.2	1.1	.9	1.7
6	4	94.2	6.9	56.7	24.4	.3	.7	1.3	1.0	.9	1.2	2.3	2.1	2.4	2.1	−.3	2.7
8	3	96.3	.4	61.1	12.0	.0	.0	.9	1.6	.9	1.6	.0	.0	.4	.8	.4	.8
9	2	89.6	14.7	62.8	26.0	.0	.0	.9	1.1	.9	1.1	1.3	1.0	2.7	1.6	1.4	2.1
3	6	91.1	17.5	66.1	35.8	1.0	1.4	2.4	1.7	1.4	1.9	3.3	1.8	5.0	.9	1.7	1.6
7	1	100.0	−	68.8	−	.0	.0	1.4	1.9	1.4	1.9	1.7	1.1	2.6	1.4	.9	1.5
5	3	97.0	5.2	73.0	10.9	.2	.7	2.2	1.6	2.0	1.7	2.3	1.5	4.9	1.5	2.6	2.7
Total	33	88.5	16.4	56.5	26.3	.2	.7	1.4	1.5	1.2	1.5	1.9	1.7	2.9	2.0	1.0	1.9

Note.—Task = compliance with task-management prompts; Word = accuracy responding to word-teaching prompts. Before refers to pretest (A phase) prior to instruction; After refers to posttest (B phase) after instruction.

Classroom-level analyses replicated these overall patterns (see Fig. 1). Children from Classroom 1 were most compliant with task-management prompts (upper right panel, Fig. 1), as compared to correct responding to word-teaching prompts. Similarly, children in Classroom 1 started low in their accuracy in responding to word-teaching prompts in Book 1, but grew thereafter across books. Similarly so, Classroom 2 children showed the same effects, but were even more stable than Classroom 1 children in complying with task-management prompts (see lower left panel, Fig. 1). They also displayed the positive, linear trend in correct responding to word-teaching prompts. Children in Classroom 3 also displayed similar trends, even though they were probed much less frequently (see lower right panel, Fig. 1) due to a combination of limited staff resources and a later start-up in the school year. Starting later in the school year, Classroom 3 children only completed instruction through Book 7. As seen in Figure 1, for Books 2, 3, and 5, their performance that was probed, performance was highly consistent with that of children in Classrooms 2 and 3 in comparable books.

Two children illustrate the two extremes in student-level responding to the narrator. Child 1 had the very lowest levels of correct responding to both task-management and word-teaching prompts, varying in the 50% to 60% range over books. In contrast, Child 3 was the most proficient responder, quickly accelerating to above 80% levels in correct task-management prompts by Book 2 and word-teaching prompts by Book 4.

Question 2: What Were the Effects on Children's Acquisition of Vocabulary Knowledge and Proficiency in Answering Story Questions?

Vocabulary. The mean pretest vocabulary score was 0.2 ($SD = 0.7$) points, increasing to 1.4 ($SD = 1.5$) for posttest, a gain of 1.2 ($SD = 1.5$) points per book. Over the series of books, an average of 9.9 points was achieved by children, ranging from 4 to 21 on the vocabulary mastery probes. This was equal to learning between five new words completely (28%) or 10 words partially (56%) out of 18 words taught in
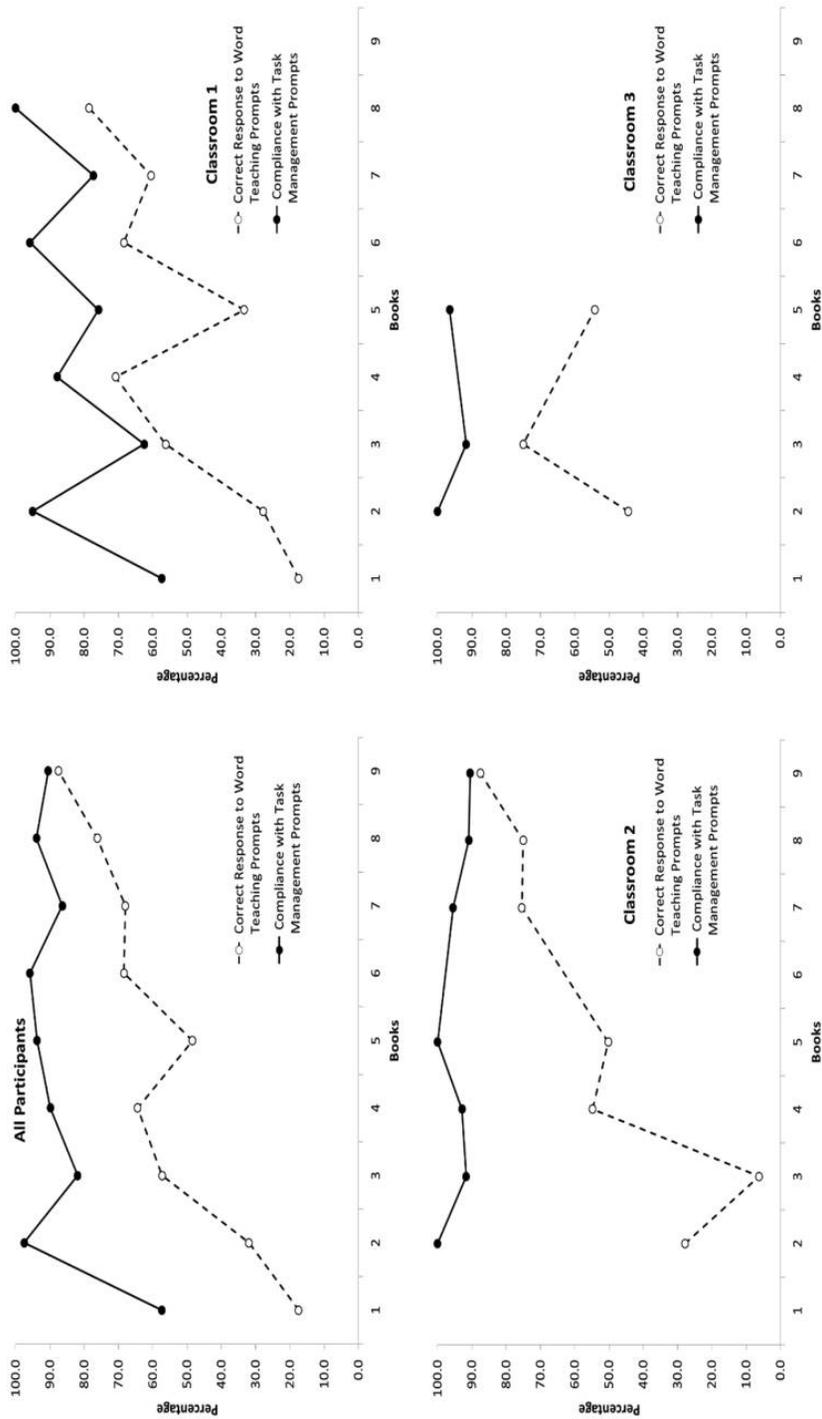
Figure 1. Mean compliance with the narrator's task-management prompts and correct responding to word-teaching prompts over books for all participants and by classroom.

the entire nine-book series. Looking at the words learned at the mastery test level, only 4% of children knew one or more words (i.e., with point scores of 2 or larger) at pretests. Following intervention, nearly half (48%) of students were learning one word or more per book ($\chi^2[1] = 36.069$, $p < .001$). These overall effects for both points and words learned were equal to an NAP effect size of 73% ($CFI_{[90\%]} = .66$ to $.80$). This represented a 46% increase in vocabulary over baseline. According to Parker et al., this represented a medium effect size for vocabulary. These results were a bit less than those reported by Spencer et al.'s (2013), at $M = 0.5$ ($SD = 0.31$) on the pretest and $M = 1.7$ ($SD = 1.4$) on the posttest (NAP = 84%).

Comprehension. Overall, story comprehension had a point score of $M = 1.9$ ($SD = 1.7$) before instruction, increasing to $M = 2.9$ ($SD = 2.0$) after, or a mean gain of 1.0 ($SD = 1.9$) per book. In terms of complete answers to questions, 52% of children (i.e., with point scores of 2 or larger) could do so at pretest. As a result of intervention, this number rose to 69% of students providing one or more complete answers ($\chi^2[1] = 5.063$, $p = .024$). These overall effects were equal to an NAP of 65% ($CFI_{[90\%]} = .58$ to $.72$). An NAP this large represented a 30% increase in comprehension after intervention compared to baseline. These values for comprehension were the same as those reported by Spencer et al. (2013) findings of $M = 2.5$ ($SD = 1.7$) before increasing to 3.6 ($SD = 1.6$), a mean gain of 1.1 after instruction (NAP = 63%). According to Parker et al., this represented a weak effect size for comprehension.

These mean gains for all children over the books series are seen in Figure 2 for vocabulary and story comprehension in the repeated-acquisition design. Consistent, repeated gains for both measures per book were seen. Like the RAD design in Spencer et al. (2013), these nine replications of acquisition across nine books
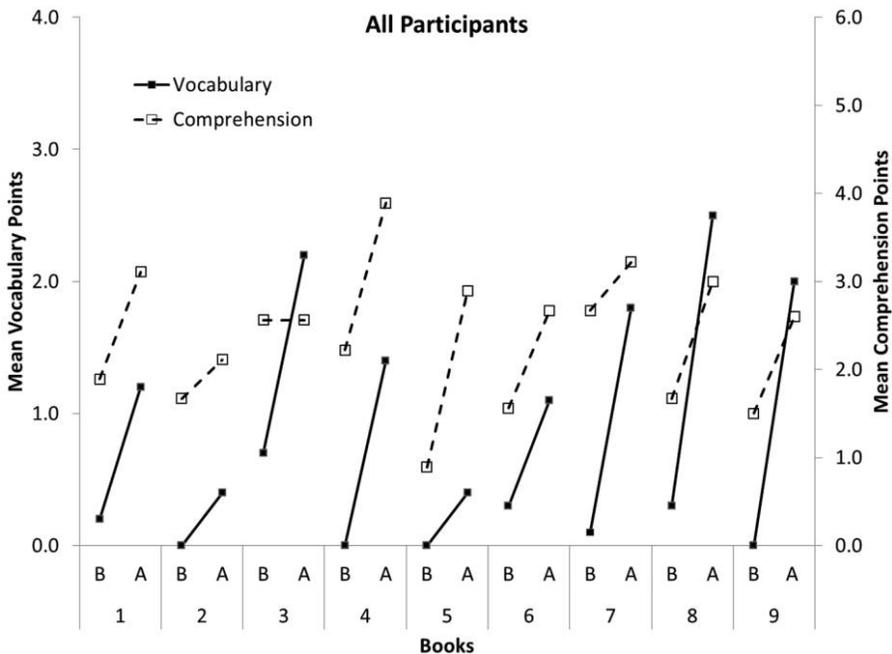

Figure 2. Repeated acquisition design plots for mean vocabulary and comprehension correct point scores over books.

in vocabulary and eight out of nine books in comprehension demonstrated experimental control of the intervention. Vocabulary gains after Books 2 and 5 were less than those of all other books. This same pattern of no gain in comprehension was also seen in Book 3 (see Fig. 2), suggesting the need for possible improvement in these books.

Individual children varied in their gains in vocabulary and comprehension over the series, as shown in Table 2. In vocabulary, the mean across individual children was 1.2, ranging from 0.4 to 2.3. Similar values for comprehension were a mean of 1.0, ranging from −0.3 to 2.6. The least responsive child (Child 1) made only two greater than zero gains out of a possible nine storybooks, or only four out of 36 possible. In contrast, Child 2 made gains in nine out of nine storybooks, earning 21 (more than half) out of a total of 36 possible vocabulary points.

Both of these children's comprehension scores were variable across books. Both at times had pretest scores larger than zero, both showed instances of no gain and negative gain, and comprehension gains did not necessarily covary systematically with vocabulary gains. However, Child 1 consistently earned more comprehension points at pretest than did Child 2. After the first two books, Child 2 compared to Child 1 thereafter demonstrated a more consistent pattern of making positive gains in comprehension points (Books 3 to 8).

Children's characteristics accounted for some variation in weekly acquisition. The overall mean vocabulary gains per book were .87 points for IEP students ($n = 3$), 1.1 for the non-IEP/ELL group ($n = 3$), and 1.6 (nearly twice as many) for the ELL group ($n = 3$). Values for comprehension points gained showed a similar rank order at 0.74, 0.96, and 1.3 for IEP, non-IEP/ELL, and ELL groups, respectively.

Question 3: Were Differences in Children's Response to the Narrator's Word-Teaching Instruction Differentially Associated with Student Learning Outcomes?

Children's compliance with the narrator's task-management prompts, $r(8) = .60$, and correct responses to narrator's word-teaching prompts, $r(8) = .50$, were both moderately and positively associated with their mean gains in vocabulary. Even though the engagement observations sampled only word-learning prompts, the correlations with comprehension gains also were positive, but weaker ($r[8] = .47$) for correct responding to word-teaching prompts but nonexistent for compliant task responding ($r[8] = .03$). Additional evidence of the positive relationships between child responding during instruction and vocabulary knowledge gain was indicated by the fact that Students 1 and 3, who were most divergent in patterns of correct responding to the narrator, also had divergent mean gains in vocabulary. For example, Child 1 had a mean gain per book of only .9, while Child 3 had a mean gain per book twice as large, at 1.9.

Discussion

The purpose of the study was to systematically replicate the promising supplemental, technology-assisted vocabulary and comprehension storybook intervention re-

ported by Spencer et al. (2013) intended for preschool children. Many vocabulary interventions have been examined with older children; the current study adds to the relatively small number of studies of vocabulary interventions for younger children. The current study added a new focus on preschool children's engagement during the intervention. We examined the hypothesis that responding to task-management prompts would be important to preschool children's learning, but that responding correctly to word-teaching prompts embedded in stories would be even more important. Several findings were noteworthy.

The first was that the same high implementation fidelity and dosage outcomes reported by Spencer et al. (2013) were replicated by different implementers in different classrooms in a different state using similar procedures. Second, child outcomes similar to those in the original study were obtained although the participant sample was quite different. Participants in Spencer et al. (2013) were African American; none had IEPs or were ELL. In contrast, our sample included three children with IEPs and three who were ELL. These findings strengthened what we know about the feasibility and efficacy of the automated intervention.

To address the first research question, we examined children's responses to task-management prompts (e.g., *turn the page* or *put your finger on the flap*) and word-learning prompts (e.g., *what does ridiculous mean?*). Results indicated that children were readily responding to task-management prompts as early as the first session. This was not the case for word-learning prompts; children started low on average and grew higher linearly over storybooks. As participants were becoming more familiar with the storybook structure, they appeared more likely to respond correctly. They also appeared to become more confident in their ability to respond correctly. These findings were replicated in Classrooms 1 and 2, where the most complete storybook data were collected and partially demonstrated in Classroom 3, which started the experiment later in the year (see Fig. 1).

We examined the effects of the intervention package on vocabulary learning and the ability to answer questions to address the second research question. Across all nine books, participants gained an average of 9.9 of a possible 36 points (28%; range 4–21) on the vocabulary mastery-monitoring probes. Because responses were scored with 2 points for a complete response and 1 point for a partial response, this increase indicates average learning of between 5 and 10 of a possible 18 words. Compared to Spencer et al. (2013), word learning effect sizes were only slightly smaller. Some children made large gains in each book compared to the very least-responsive child, who made vocabulary gains in only two of the nine books.

The children with IEPs and ELL were the most challenged by their lower total language fundamental scores at start (see Table 1). Even so, as a group the three ELL children made greater gains per book in both vocabulary knowledge and comprehension than did children with IEPs or the other children (neither ELL nor an IEP). This suggested that the oral language instruction in English was relatively successful. However, these findings are preliminary at best, and generality to other samples with these characteristics is uncertain without future replication.

Overall, these word-learning findings were strong considering the relatively brief intervention (approximately 7 hours) and the weak preintervention skills of participants. Also, the definitional word task was a rigorous test of word learning: preschool children were asked to generate a decontextualized definition

or synonym to indicate learning of the word. Studies of vocabulary interventions that have selected similarly challenging words taught with similarly explicit approaches typically show that children learn only a proportion of the words taught. Beck and McKeown (2007) reported that 33%–39% of total words taught were learned under conditions of extended, explicit instruction. Loftus, Coyne, McCoach, and Zipoli (2010) reported posttest scores of 36% on the six words taught. Pullen et al. (2010) reported gains of 27% using a second tier of supplemental vocabulary instruction to a group of students with a mean PPVT standard score of 74.80 ($SD = 13.67$). Taken in context, the words learned by children in this automated intervention compared favorably to these previous reports.

Children gained an average of 1.0 comprehension point out of a possible six points per book, or less than one story question per book, slightly less than reported by Spencer et al. (2013). One challenge to the assessment of the story questions was that many children could answer story questions at pretest (prior to having heard the story). Fifty-two percent of children answered one or more complete questions. Although this number increased to 69% of children answering one or more complete story questions at posttest, this was a weak effect size, according to Parker and Vannest (2009). Compared to new-word learning, it appeared that the intervention was not nearly as successful in creating changes in participants' ability to answer story questions.

Our third research question addressed the association between participants' responses to word-learning prompts and subsequent word learning. As hypothesized, there was a moderate, positive relationship between children's frequency of correct responding to the narrator's word-teaching prompts and their new-word-learning outcomes. The same relationship between task-management prompts and new-word learning, however, was also moderately large. These correlational relationships did not support our differential relationship hypothesis; instead, both types of engagement were important for learning outcomes. The finding supports future experimental examination of these relationships.

Implications for Use of Technology

The intervention in its current form consisted of storybooks and accompanying prerecorded audio that children in small groups heard through headphones to receive supplemental instruction facilitated by an aide that did not require the teacher's participation. This automated listening station approach used an ordinary device (i.e., MP3 player with splitter for multiple listeners) for audio playback and was relatively cost efficient, reliable, available, and usable compared to other electronic media and devices on the market that might serve this purpose. Results indicated that children were relatively successful in following and responding to the prompts provided by the narrator, especially the task-management prompts, and later the word-learning prompts as well. Modifications to this intervention approach are worth exploring. One possibility is that preschool children might be able to use the listening center without the assistance of the adult facilitator. Our results suggest that participants responded correctly to task-management

prompts after experience with the first book. Perhaps initial support from the adult facilitator could be faded out as participants demonstrated independence. Another possibility would be the delivery of the same or similar storybook content on iPads or other tablet devices. As in the current delivery, small groups of children using tablet devices may require or benefit from monitoring by a facilitator to ensure that individuals are responding to the narrative and using their book appropriately. Streaming the content to children through the internet simultaneously to all devices would enable synchronization of all children and the facilitator. It also is possible that the current approach could be developed to support core (Tier 1) universal instruction, wherein each child has a storybook but the narration is provided over the classroom PA or digital TV systems to make the audio publicly audible.

Limitations

Limited resources necessitated that research staff serve multiple roles (i.e., observers, testers, and intervention facilitators). With purposeful scheduling, we achieved a mix of children, classrooms, and observers/facilitators to limit bias because observers were not always facilitators or testers of the same children. Another concern was facilitators responding as participants during lessons, and whether or not this added any unique prompting and/or response modeling/feedback that were outside of the planned intervention. Mitigating this concern was explicit training of the facilitator to not provide instruction, and fidelity scores were high. Additionally, the facilitator's role was considered a part of the intervention's package of components. It was not expected that children would learn uniquely from the technology in the absence of a small-group facilitator. Interpreting findings of the full intervention package was not confounded.

Resource restrictions also precluded complete observation of students' responding to the narrator's word-teaching prompts in all books and classrooms. Eight of the nine books were observed in two classroom groups, and only three of nine books in the third. Across children, the number of probes ranged from one to six. Because some children were sampled more frequently than others, there was a possibility that participant variability in the data plotted in Figure 1 may have influenced the results. However, multiple children were observed for each book, with some variation in order and occasion, thus lessening bias. Limited resources also restricted tracking of children's engagement to only word-teaching but not the comprehension-teaching prompts.

Research assistants rather than classroom staff were the implementers of the storybook intervention, and results were those observed under high fidelity and planned dosage. We had the luxury of providing make-up sessions for children with absences who missed weekly sessions. Similar luxuries in naturalistic settings may not exist. The mastery-monitoring probes used in the current study provided a proximal, immediate measure of intervention effects. Additional information could be obtained from measures of retention of knowledge, or by the inclusion of a generalized measure of children's vocabulary knowledge as an outcome. Although few studies of targeted interventions have reported significant effects on

generalized vocabulary measures, the lack of a generalized measure in the current study limits the interpretation of findings.

Implications for Future Research and Practice

Evidence of internal validity challenges to the RAD (i.e., equivalence and sequence) was not seen. Children learned measurably more after experiencing book listens. The least growth occurred with storybooks 2 and 5 compared to the other storybooks; however, these differences were probably more related to seasonal issues (i.e., the holidays) than variance in word equivalence. Similarly, there was not linear growth over books suggestive of order effects. Future studies can easily alter the sequence of storybooks to better examine this issue in RADs.

Given the findings from the first study and this replication, both with small samples, evaluating the efficacy in a larger randomized control trial with preschool staff implementing/facilitating the technology-based intervention is warranted. Future research is needed to establish whether acquisition of a small number of words is likely to impact later reading achievement. It is possible that early improvements in oral language skill, including vocabulary, will help prepare students to take advantage of later language-learning opportunities (Penno et al., 2002; Robbins & Ehri, 1994). Experience with new-word learning may help children develop metalinguistic skills that will facilitate future word learning. Areas of possible intervention improvements for future research emerged. Such experience is also beneficial to dual language learners (Gutiérrez, Zepeda, & Castro, 2010).

The learning curve identified in children's accurate responding to the narrator's word-teaching prompts over storybooks suggests a need for more work in this automated format to accelerate children's initial correct responding and sustain it over time. Only one lesson was used to orient and teach the children how to learn in the automated program. A stronger effort at initial training—for example, using several rather than only one lesson prior to the first storybook—might be an improvement. Similarly, working individually with the least responsive children in their initial and subsequent training might also work.

Children were much more successful in learning new words than in answering questions about the stories. This was also the case in Spencer et al. (2013). In the current study, we did administer the comprehension pretests prior to the children having any listens to the storybook. The data indicated that 52% of children could answer at least one question at pretest, apparently by guessing rather than by reporting information unique to the story. For example, *In this story, Ellie is happy. Why is Ellie happy?* Answer: *Because she likes playing?* This was a correct answer before having heard the story. This was most likely a good guess drawing on the child's prior experience with play and happiness. A first challenge to the development of an effective comprehension intervention is the identification of appropriate instructional targets. As indicated by the previous example, there is a need for targets that are sufficiently difficult to reduce the likelihood of guessing correctly, but not too difficult for preschool children to answer after instruction. A second challenge is the use of an automated format for comprehension intervention. In its current form, the automated intervention may contain too few response trials (dosage), and contingent feedback (not available in this format) may be required to

do a better job promoting story comprehension. Third, future improvements in the comprehension measure could benefit from an in-depth analysis of what the children actually say in their answers to storybook questions, particularly content they generate beyond what they were actually taught.

Implications for Practice

Preschools with children needing MTSS supports for children should consider using this Tier 2 intervention, particularly given the findings reported in this second replication study. This automated intervention would provide vocabulary instruction to children with weak skills to supplement the high-quality core language instruction delivered in Tier 1. The words taught in Tier 2 are sophisticated vocabulary with potential for use in a variety of language contexts. Delivery via the use of technology is efficient, supports high fidelity of intervention delivery, and requires minimal teacher participation.

In an MTSS model, vocabulary development can be supported in each tier. Tier 1 instruction might well emphasize brief instruction and repeated exposure to words that could be acquired in that manner. This Tier 2 intervention is more intensive, providing explicit instruction to the teaching of challenging words. Tier 3 interventions can target the needs of children with very weak skills with a focus on more basic vocabulary, presented one-on-one by the teacher using highly scripted, explicit instruction with many opportunities to respond and receive feedback (Kaminski, Powell-Smith, Hommel, McMahon, & Bravo-Aguayo, 2015). In combination, these language and literacy support options could produce improvements.

Conclusion

A promising Tier 2 preschool supplemental, technology-assisted storybook oral language intervention delivered in small groups was replicated in a small but diverse sample of preschoolers with identified weak oral language skills. These new findings broadened the support for the initial findings of Spencer et al. (2013), and extended them in several areas (e.g., engagement with the narrator, different learners, etc.). Demonstration that similar findings can be produced using the intervention when classroom paraprofessionals or volunteers are the implementers is the logical next step in future research developing this intervention for MTSS.

Note

References

Barnett, W. S., Epstein, D. J., Friedman, A. H., Sansanelli, R., & Hustedt, J. T. (2011). *The state of preschool 2009*. National Institute of Early Education Research, Rutgers. Retrieved from http://nieer.org/yearbook/pdf/yearbook.pdf

Beck, I. L., & McKeown, M. (2007). Increasing young children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal*, **107**, 251–272.

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. **2**, pp. 41–51). New York: Guilford.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology*, **98**(1), 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.

Bradfield, T., Besner, A., Wackerle-Hollman, A., Albano, A., Rodriguez, M., & McConnell, S. (2013). Redefining individual growth and development indicators: Oral language. *Assessment for Effective Intervention*, **39**(4), 233–244.

Buysee, V., & Peisner-Feinberg, E. S. (2013). *Handbook of response to intervention in early childhood*. Baltimore, MD: Brookes.

Catts, H. W., Fey, M., Zhang, X., & Tomblin, J. B. (2001). Estimating risk for future reading difficulties in kindergarten children: A research-based model and its clinical implications. *Language, Speech, and Hearing Services in Schools*, **32**, 38–50.

Chambers, B., Slavin, R. E., & Madden, N. (2011). Small-group, computer-assisted tutoring to improve reading outcomes for struggling first and second graders. *Elementary School Journal*, **111**(4), 625–640.

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

Dickinson, D. K. (2011). Teachers' language practices and academic outcomes of preschool children. *Science*, **333**, 964–967.

Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effect of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly*, **22**(2), 243–260.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test—Fourth Edition*. Bloomington, MN: Pearson Assessments.

ECO Center. (2011). *SPR reporting requirements*. Washington, DC: U.S. Department of Education. Retrieved from http://www.ed.gov/about/reports/annual/2007plan/program.html

Gettinger, M., & Stoiber, K. (2007). Applying a response to intervention process for early literacy development in low-income children. *Topics in Early Childhood Special Education*, **27**, 198–213.

Gibson, L., Cartledge, G., & Keyes, S. E. (2011). A preliminary investigation of supplemental computer-assisted reading instruction on the oral reading fluency and comprehension of first-grade African American urban students. *Journal of Behavioral Education*, **20**(4), 260–282.

Gillam, R. B., Loeb, D. F., Hoffman, L. M., Bohman, T., Champlin, C. A., Thibodeau, L., & Friel-Patti, S. (2008). The efficacy of Fast ForWord language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language and Hearing Research*, **51**(1), 97–119.

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

Greenwood, C. R., Carta, J. J., Goldstein, H., Kaminski, R., McConnell, S., & Atwater, J. (2014). The Center on Response to Intervention in Early Childhood (CRTIEC): Developing evidence-based tools for a multi-tier approach to preschool language and early literacy instruction. *Journal of Early Intervention*, **36**(4), 246–262.

Greenwood, C. R., & Kim, J. M. (2012). Response to Intervention (RTI) services: An eco-behavioral perspective. *Journal of Education and Psychology Consultation*, **22**, 1–27.

Gutiérrez, K. D., Zepeda, M., & Castro, D. C. (2010). Advancing early literacy learning for all children: Implications of the NELP report for dual-language learners. *Educational Researcher*, **39**(4), 334–339.

Head Start. (2010). *The Head Start child development and early learning framework promoting positive outcomes in early childhood programs serving children 3–5 years old: Revised*. Washington, DC: Administration for Children and Families. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/eecd/Assessment/Child%20Outcomes/HS_Revised_Child_Outcomes_Framework.pdf

Hindman, A., Connor, C., Jewkes, A., & Morrison, F. (2008). Untangling the effects of shared book reading: Multiple factors and their associations with preschool literacy outcomes. *Early Childhood Research Quarterly*, **23**, 330–350.

Horner, R. H., Carr, E. G., Halle, J. W., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, **71**, 128–148.

Johnston, P. H. (2011). Response to intervention in literacy. *Elementary School Journal*, **111**(4), 511–534.

Justice, L. M., Hamre, B. K., & Pianta, R. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, **23**, 51–68.

Justice, L. M., 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**(4), 983–1001.

Kaminski, R. A., Powell-Smith, K., Hommel, A., McMahon, R., & Bravo Aguayo, K. (2015). Development of a Tier 3 curriculum to teach early literacy skills. *Journal of Early Intervention*, **36**, 313–332.

Kelley, E. S., & Goldstein, H. (2014). Building a Tier 2 intervention: A glimpse behind the data. *Journal of Early Intervention*, **36**(4), 292–312

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

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. What Works Clearing House. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Landry, S. H., Assel, M. A., Anthony, J. L., & Swank, P. R. (2013). Developing a universal screening and progress monitoring tool and its applicability for us in response to intervention. In V. Buysee & E. S. Peisner-Feinbert (Eds.), *Handbook of response to intervention in early childhood* (pp. 155–168). Baltimore, MD: Brookes.

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.

Lonigan, C. J., 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.

Macaruso, P., & Walker, A. (2008). The efficacy of computer-assisted instruction for advancing literacy skills in kindergarten children. *Reading Psychology*, **29**(3), 266–287.

National Center for Special Education Research (NCSER). (2014). Request for applications: Special education research grants—FY 2015 (CFDA No. 84.332A). Retrieved from http://ies.ed.gov/funding/pdf/2015_84324A.pdf

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, **40**, 357–367.

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.

Powell, D. R., Burchinal, M., File, N., & Kontos, S. (2008). An eco-behavioral analysis of children's engagement in urban public school preschool classrooms. *Early Childhood Research Quarterly*, **23**, 108–123.

Pullen, P. C., & Justice, L. M. (2003). Enhancing phonological awareness, print awareness, and oral language skills in preschool. *Intervention in School and Clinic*, **39**, 87–98.

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.

Riley-Tillman, T. C., & Burns, M. K. (2009). *Evaluating educational interventions: Single-case design for measuring response to intervention*. New York: Guilford.

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology*, **86**, 54–64.

Savage, R. S., Abrami, P., Hipps, G., & Deault, L. (2009). A randomized controlled trial study of the ABRACADABRA reading intervention program in grade 1. *Journal of Educational Psychology*, **10**(3), 590–604.

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.

Shanahan, T., & Lonigan, C. J. (2008). *Developing early literacy: A report of the National Early Literacy Panel*. Retrieved from http://www.nifl.gov/publications/pdf/NELPReport09.pdf

Spencer, E. J., Goldstein, H., & Kaminski, R. (2012). Teaching vocabulary in storybooks: Embedding explicit vocabulary instruction for young children. *Young Exceptional Children*, **15**(1), 18–32.

Spencer, E. J., Goldstein, H., Sherman, A., Noe, S., Tabbah, R., Ziolkowski, R., & Schneider, N. (2013). Effects of an automated vocabulary and comprehension intervention: An early efficacy study. *Journal of Early Intervention*, **14**(1), 195–221.

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). New York: Springer.

van Kleeck, A. (2006). Fostering inferential language during book sharing with prereaders: 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: Plural Publishing.

van Kleeck, A. (2008). Providing preschool foundations for later reading comprehension: The importance of and ideas for targeting inferencing in storybook-sharing interventions. *Psychology in the Schools*, **45**, 627–643.

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.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, **45**, 225–256.

Walsh, B., & Blewitt, P. (2006). The effect of questioning style during storybook reading on novel vocabulary acquisition of preschoolers. *Early Childhood Education Journal*, **33**, 273–278.

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology*, **98**(1), 63–74.

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

Wild, M. (2009). Using computer-aided instruction to support the systematic practice of phonological skills in beginning readers. *Journal of Research in Reading*, **32**(4), 413–432.

Williford, A. P., Vick Whittaker, J. E., Vitiello, V. E., & Downer, J. (2013). Children's engagement in preschool and the development of self-regulation. *Early Education and Development*, **24** (2), 162–187.