JSLHR

Research Article

Academic Vocabulary Learning in First Through Third Grade in Low-Income Schools: Effects of Automated Supplemental Instruction

Howard Goldstein,[a] Robyn A. Ziolkowski,[b] Kathryn E. Bojczyk,[c] Ana Marty,[d] Naomi Schneider,[e] Jayme Harpring,[e] and Christa D. Haring[a]

Purpose: This study investigated cumulative effects of language learning, specifically whether prior vocabulary knowledge or special education status moderated the effects of academic vocabulary instruction in high-poverty schools.
Method: Effects of a supplemental intervention targeting academic vocabulary in first through third grades were evaluated with 241 students (6–9 years old) from low-income families, 48% of whom were retained for the 3-year study duration. Students were randomly assigned to vocabulary instruction or comparison groups.
Results: Curriculum-based measures of word recognition, receptive identification, expressive labeling, and decontextualized definitions showed large effects for multiple levels of word learning. Hierarchical linear modeling revealed that students with higher initial Peabody Picture Vocabulary Test–Fourth Edition scores (Dunn & Dunn, 2007) demonstrated greater word learning, whereas students with special needs demonstrated less growth in vocabulary.
Conclusion: This model of vocabulary instruction can be applied efficiently in high-poverty schools through an automated, easily implemented adjunct to reading instruction in the early grades and holds promise for reducing gaps in vocabulary development.

Researchers have found that children who begin elementary school with deficits in their vocabulary knowledge are at risk for reading failure and academic underachievement (e.g., Catts, Adlof, & Weismer, 2006; Marzano, 2003; Nagy, 2005). According to the National Reading Panel (NRP; 2000), vocabulary serves as a pivotal link between oral and written language. In early literacy development, decoding is heavily dependent on the words in students' vocabulary (Duff & Hulme, 2012; Ehri, 1998). If a decoded word is not present in a student's repertoire, then the beginning reader will not be able to discern its meaning from print (Pikulski & Chard, 2005). Thus, students with limited vocabularies are at risk for difficulties in developing fluent reading skills and in comprehending reading passages (Nash & Donaldson, 2005; Scarborough, 2005). For children who are at risk for reading disabilities due to limited vocabulary, explicit teaching may be particularly important (Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009; Foorman & Torgesen, 2001). To determine the potential for overcoming limited vocabularies, there is a need to examine the effects of explicit vocabulary instruction for children identified with developmental delays as well as children at risk for language and reading disabilities.

Vocabulary Development

Before entering school and learning how to read, children learn vocabulary from everyday interactions with parents and caregivers (Hoff, 2003; Snow, 1994). Word learning begins when infants isolate and store sound segments garnered from the sound stream of their native language (Saffran, Werker, & Werner, 2006). At approximately 6 months of age, infants can recognize frequently heard words, for example, their own names (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005), and by 10–12 months of age, children are linking words to objects or concepts to which they have frequent exposure (Nelson, 1973; Phythian-Sence & Wagner, 2007). By 30 months, estimates for expressive vocabularies range from 360 to 630 words, and estimates

[a]University of South Florida, Tampa
[b]University of Northern Colorado, Greeley
[c]The Catholic University of America, Washington, DC
[d]Florida State University, Tallahassee
[e]The Ohio State University, Columbus
Correspondence to Howard Goldstein: hgoldstein@usf.edu

Editor-in-Chief: Sean Redmond
Editor: Geralyn Timler
Received March 16, 2017
Revision received June 10, 2017
Accepted June 29, 2017
https://doi.org/10.1044/2017_JSLHR-L-17-0100

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

or receptive vocabularies are even larger (Fenson et al., 1994; Tomasello, 2003). But Rowe, Raudenbush, and Goldin-Meadow (2012) found that children with lower vocabularies at 30 months demonstrated reduced vocabulary growth over time, thus jeopardizing their school readiness. In fact, researchers have demonstrated that the size of children's receptive and expressive vocabulary in the second year of life predicts later vocabulary, phonological awareness, reading accuracy, and reading comprehension (Duff, Reen, Plunkett, & Nation, 2015; Lee, 2011).

Because vocabulary growth is not an all or none phenomenon, it is difficult to characterize vocabulary learning. Children are thought to learn words incrementally through multiple exposures in various contexts (Anderson & Nagy, 1991; Beck, McKeown, & Omanson, 1987; Dale, 1965; Hatch & Brown, 1995; Stahl, 1986). This learning process facilitates deeper word knowledge than a single exposure (Bolger, Balass, Landen, & Perfetti, 2008). Investigators have indicated that preschoolers (Hadley, Dickinson, Hirsch-Pasek, Golinkoff, & Nesbitt, 2015) and kindergarteners (Christ, 2011) go through several stages when learning new words. For example, Dale (1965) outlined a typical progression for learning novel vocabulary words. These stages are as follows: Level 1, "never heard it before"; Level 2, "heard it, but do not know what it means"; Level 3, "recognize it in context as having something to do with____"; Level 4, "know it well." Not until knowledge of a given word reaches Level 3 is the child able to identify it, use it, and/or provide attributes of it within a context. Once Level 4 is reached, the child can define the word, use it in more than one context, and provide salient attributes without a defined context. Children who have more words in their vocabulary at Levels 3 and 4 will demonstrate better reading comprehension and better use of context clues when novel words are encountered (NRP, 2000). These developmental levels seem to correspond to depth of knowledge. When vocabulary is being taught explicitly, researchers can use multiple measures of word learning to begin to determine if the developmental process progresses as hypothesized by Dale and others (Beck et al., 1987; Hatch & Brown, 1995; Stahl, 1986).

There is considerable variability in the rate that words are acquired (Beck & McKeown, 1991). Researchers have attributed these differences to environmental factors such as family socioeconomic status (SES), maternal gestures, the amount or frequency of linguistic input and exposure to diverse vocabulary words in the child's early environment, and early literacy practices such as shared storybook reading (Dickenson & Porche, 2011; Hart & Risley, 1995; Hoff, 2003; Pan, Rowe, Singer, & Snow, 2005; Payne, Whitehurst, & Angell, 1994; Raikes et al., 2006; Rowe et al., 2012). Hart and Risley's (1995) landmark study found that, long before formal schooling, vast differences in vocabulary exposure typify the language experiences of children from poverty when compared to working and professional class homes. Hart and Risley (1995) estimated that children who come from low-income homes know approximately 6,000 fewer words than their middle-class peers when they enter school. These initial differences in vocabulary knowledge have been shown to increase over time, widening the gap between students with more sophisticated vocabularies and those with depressed vocabularies (Duff et al., 2015; Stanovich, 1986; Walker, Greenwood, Hart, & Carta, 1994).

What Words to Teach

No research-based formula currently exists for choosing the most functional and meaningful vocabulary words to teach, despite the numerous word lists that exist at different grade levels (Baumann & Kame'enui, 2004; Biemiller, 2005). However, recent education policies seem to recognize the need for proficient academic oral language to enable students to access content in academic texts and to develop high-level academic discourse (Bailey & Heritage, 2008). For example, the majority of states have elected to adopt the Common Core State Standards (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). The Language Arts and Literacy strands contain clear guidance about the hierarchical grade-level skills in reading, writing, speaking, listening, and language that students are expected to master from kindergarten to 12th grade. One important component in the Language Arts strand is academic vocabulary. Academic vocabulary refers to words that occur frequently across the curriculum in spoken or written academic discourse (Baumann & Graves, 2010; Coxhead, 2000; Marzano & Pickering, 2005). These words are fundamental for text comprehension (Beck & McKeown, 2007). They are considered to have a high probability of not being known by students (Hiebert & Lubliner, 2008). Precise meanings of academic words vary based on context, for example, *operation, specific, investigate* (Beck, McKeown, & Kucan, 2002; Marzano & Pickering, 2005). Learning academic vocabulary words and understanding their specialized meanings may be particularly challenging for children who lack a rich foundational vocabulary (Stahl & Nagy, 2006). Direct instruction on academic vocabulary words beyond what is targeted in the general curriculum may be the only means for children at risk to increase the breadth and depth of their vocabularies needed to boost their academic achievement.

Beck et al. (2002) proposed a tiered system to prioritize vocabulary for instruction. Tier 1 contains basic words that rarely require direct instruction (e.g., *boy*); Tier 3 words are situation specific and are seldom used outside of a specific context (e.g., *decibel*). In contrast, Tier 2 words are found across a variety of linguistic and academic environments (e.g., *layer, consequence, adult*). Academic vocabulary incorporates both Tier 2 and Tier 3 and is of considerable importance for developing the academic language needed for children to be successful in many different content areas (Nagy & Townsend, 2012). However, Beck et al. (2002) suggest prioritizing Tier 2 words to maximize the impact on oral and reading vocabulary knowledge of children who are at risk for developing reading disabilities. Hence, teaching

academic vocabulary that qualifies as Tier 2 words have the greatest potential long-term benefit.

How to Teach Vocabulary

Given the disparities in children's vocabularies at school entry, one would expect that enhancing children's vocabulary would be a high priority from the beginning of primary school (Biemiller & Boote, 2006; Graves, 2016; National Early Literacy Panel, 2008). That is not the case, as investigators have found low rates of vocabulary instruction in elementary schools (Biemiller, 2004; Scott, Jamieson-Noel, & Asselin, 2003; Wright & Neuman, 2014). Moreover, there is no established method for teaching vocabulary in the primary grades (NRP, 2000). The research literature on vocabulary intervention seems to overlook the early primary grades as well, as most studies on vocabulary instruction have enrolled students between third and eighth grade (NRP, 2000). Thus, key strategies must be gleaned from research conducted with children in third grade and above. Explicit vocabulary instruction that connects new vocabulary to prior knowledge is thought to have the largest effect on vocabulary gains (Baumann & Kame'enui, 2004). Beck et al. (2002) outlined principles to incorporate into instruction to enhance student learning. These principles include directly explaining words in a child-friendly format, encouraging active engagement in learning, presenting the words in rich language contexts with multiple examples, and restructuring the words or tasks when needed. Others have incorporated other features, such as opportunities for extended talk (Neuman & Wright, 2013), multimedia instruction (Neuman, 2013), or multidimensional instruction (Silverman, 2007). A critical component of instruction is to ensure that children have lots of opportunities to respond (Greenwood, Horton, & Utley, 2002); this must be accomplished efficiently and equitably. A conceptual model of change summarizing our logic behind explicit vocabulary instruction is depicted graphically in Figure 1.

These instructional strategies from upper-level grades have been adapted to studies conducted with students in kindergarten through third grade with positive effects on word learning. For example, many investigators have embedded vocabulary lessons in shared book reading for students in the primary grades, (Biemiller & Boote, 2006; Coyne, McCoach, & Kapp, 2007; Maynard, Pullen, & Coyne, 2010; Penno, Wilkinson, & Moore, 2002; Zipoli, Coyne, & McCoach, 2011). These embedded lessons include brief, child-friendly explanations of the meaning of target vocabulary words incorporated into repeated readings of story books. In some cases, extended instruction has been provided that included both embedded lessons and exposure to the words in contexts outside the story (Beck & McKeown, 2007; Coyne et al., 2009; Puhalla, 2011; Zipoli et al., 2011). When novel vocabulary is introduced during shared book reading, children have benefited from explicit vocabulary instruction (Biemiller & Boote, 2006; Justice, Meier, & Walpole, 2005; Kelley & Goldstein, 2014; Kelley, Goldstein, Spencer, & Sherman, 2015; Loftus, Coyne, McCoach, Zipoli, & Pullen, 2010; Marulis & Neuman, 2010; Sénéchal & Cornell, 1993; Spencer et al., 2012).

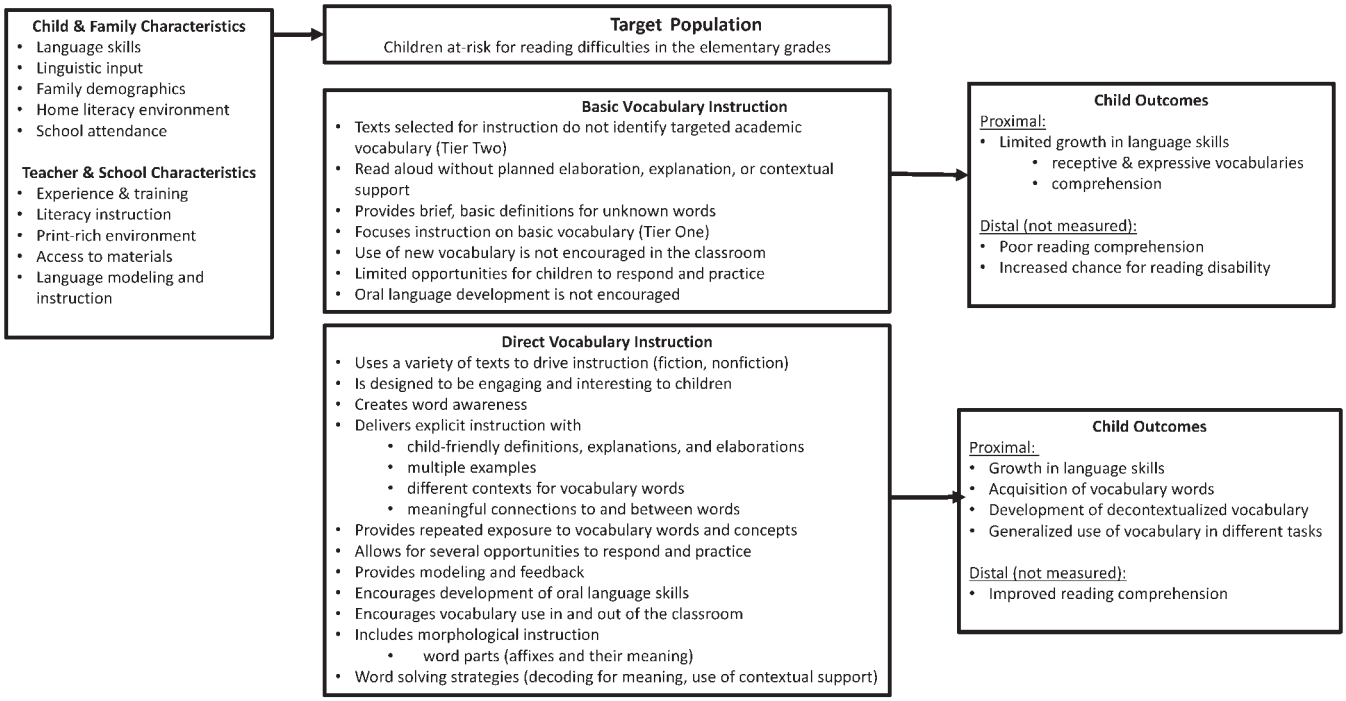
Investigators also have incorporated these instructional methods into more direct approaches to teaching vocabulary in the primary grades that did not include a shared storybook component (Apthorp, 2006; Nash & Snowling, 2006). For example, Nash and Snowling (2006) investigated the effects of teaching definitions to teaching context clues using semantic maps and extended discussion of words for children between the ages of 7;6 to 8;8 years with poor vocabulary knowledge. Students in the experimental group were taught to use contextual cues with semantic maps (i.e., Venn diagrams) to increase their understanding of word definitions. Results of the intervention, provided twice a week for 30 min over 6 weeks, showed that the students receiving the context clue/semantic mapping instruction scored higher on maintenance tests of semantic knowledge.

Cumulative Effects of Vocabulary Learning

Contradictory findings call into question the role of prior vocabulary knowledge on the learning of novel words. Some investigators have found that children who learn the most words through explicit vocabulary instruction generally begin with higher initial vocabulary knowledge (Coyne, Simmons, Kame'enui, & Stoolmiller, 2004; Penno et al., 2002; Robbins & Ehri, 1994). This has been described in the research as the "Matthew effect" (Cain & Oakhill, 2011; Stanovich, 1986), which predicts that children with richer vocabularies get richer and those with poorer vocabularies fall further behind. Because new vocabulary knowledge builds on existing knowledge, children who are linguistically disadvantaged continue to remain behind and will likely require instruction that goes beyond general classroom exposure over an extended time to demonstrate gains similar to linguistically advantaged peers.

Whether the Matthew effect applies to vocabulary learning and whether a hypothesized moderating role of initial vocabulary status exists remain unclear. In their meta-analysis of preschool vocabulary instruction, Marulis and Neuman (2010) found no difference between gains on vocabulary measures for children at risk and all other children. However, when low SES status was combined with at least one other risk factor (e.g., second language status, low academic achievement, special education status), differential effects were significant. Justice, Meier, and Walpole (2005) found that kindergarten children with lower initial vocabulary scores showed more growth in Tier 2 vocabulary learning after receiving 20 small group storybook reading sessions over 10 weeks. These findings corroborated those of Elley (1989), who determined that children with the lowest baseline vocabulary ability made greater gains over the high baseline ability group after receiving a word-learning intervention that utilized repeated book readings. In contrast, other investigators have found that children with lower vocabulary scores made smaller gains in learning (Coyne et al., 2004, 2009; Penno et al., 2002;

Figure 1. Visual depiction of theory of change motivating explicit vocabulary instruction.



Pullen, Tuckwiller, Konold, Maynard, & Coyne, 2010). For example, Penno et al. (2002) found that children with higher vocabulary ability scored higher on receptive and expressive outcome measures than children with lower ability after receiving a storybook intervention that included definitions of vocabulary words, role plays, and highlighted illustrations of words. Similarly, Coyne et al. (2007) found that, although the most at-risk students experienced significant word learning with extended vocabulary instruction, their response was below that of students who were determined to be at less risk.

Effects of Disability Status on Vocabulary Learning

In addition to at-risk populations, it is important to study the effects of disability status on the effects of vocabulary instruction. Kan and Windsor (2010) conducted a meta-analysis of word-learning instruction with children with language impairments (LIs). Results indicated that children with LI had significantly lower novel word learning than typically developing age-matched peers. The mean effect size between the LI group and a group matched for age was a 0.6 standard deviation difference. They also found that (a) verbs were harder to learn than nouns; (b) children with LI needed more exposure than typically developing children to learn words; and (c) in all novel word-learning tasks, children with typical language abilities outperformed children with LI. Similarly, researchers have found that children with learning disabilities lack strategies to learn words from context (Steele & Watkins, 2010), have less complete knowledge of words (Swanson, 1986), and do not engage in independent reading at a level necessary to significantly improve vocabulary development (Baker, Simmons, & Kame'enui, 1998). Nevertheless, Cirrin and Gilliam (2008) found clear benefits of intervention in their systematic review of language intervention for school-age children with language disorders. Their analyses indicated that direct instruction, interactive conversational reading strategies, and slowed presentation of stimuli, even in larger groups, were efficacious for improving semantic processing and vocabulary in children with LI.

Supplemental Vocabulary Interventions

Few studies have investigated supplemental vocabulary interventions implemented in the primary grades (e.g., Apthorp, 2006; Loftus et al., 2010; Puhalla, 2011). The goal in developing supplemental programming is to provide additional vocabulary instruction in an attempt to close the ever widening early vocabulary gap in the primary grades (Walker, Greenwood, Hart, & Carta, 1994). For example, Puhalla (2011) provided 22 randomly assigned first-grade students supplemental explicit instruction on 19 storybook vocabulary words extracted from the classroom-based *Read Aloud* curriculum. Students receiving scripted, 20-min lessons with explicit definitions and extension activities such as word play on the targeted vocabulary scored higher on a researcher-developed curriculum-based assessment, when compared to a no-treatment control. Medium to large effect sizes were noted, but the investigators did not provide specific results of word learning for the targeted 19 words.

In another supplemental program, Apthorp (2006) utilized a published vocabulary guide, *Elements of Reading (EOR): Vocabulary, Level C*. Teachers used a scripted guide and student book to deliver 20 min of daily whole-classroom supplemental instruction that focused on seven words per week to third-grade students in seven Title I schools across two states for 24 weeks (i.e., a total of 168 words). Fifteen teachers were randomly assigned to use the intervention or continue to provide instruction as usual. Instead of assessing word knowledge on the 168 words targeted in the intervention, the investigator utilized a standardized test of oral and sight vocabulary. The supplemental instruction teachers produced significantly better effects in the lower-income state, but not the other state. Even with improved outcomes, the students in the treatment classrooms did not reach grade-level expectations on tests of reading vocabulary and comprehension.

Loftus et al. (2010) investigated supplementary vocabulary interventions via shared storybook reading for 20 at-risk kindergarten students who participated in a whole-classroom shared story book reading instructional component supplemented by a 2-week, 30-min intervention that focused on (a) a review of the target word meanings, (b) student elaboration and extension of vocabulary word learning, and (c) individual feedback provided by graduate student interventionists. Students in the experimental group received 4 hr of instruction on eight target vocabulary words. Researchers assessed word meanings based on a continuum of partial word knowledge to complete knowledge. Students who received the supplemental instruction significantly increased their knowledge of vocabulary as evidenced by researcher-made measures of word recognition (86%), context questions (26%), and expressive definitions (28%). Effect sizes ranged from 0.42 to 0.69. Word learning in the at-risk experimental group approached those of their not at-risk peers who received classroom instruction but did not match those gains.

Although these short-term interventions have produced positive word learning outcomes, they do not show evidence of longitudinal learning sufficient to close the vocabulary gap and reduce the Matthew effect for children at risk for and with disabilities. The current study examines the effects of explicit vocabulary instruction delivered longitudinally to children in low-income schools in the early primary grades. The study was designed to examine effects for up to 3 years of explicit academic vocabulary instruction. To ensure that children were given ample and equitable opportunities to respond to academic vocabulary instruction, prerecorded lessons were embedded within stories as a supplement to reading instruction and delivered four times per week. Lessons were consistent with Beck et al.'s (2002) principles of robust vocabulary instruction. As children listened to weekly stories, worksheets provided visual information that corresponded to the words being introduced within the story and opportunities to respond. Multiple measures of vocabulary learning were administered to track the levels of word learning of up to 324 words after four units of instruction per year in first through third grade. The potential moderating effects of baseline vocabulary levels and disability status were included to determine whether these factors affected learning.

Our chief aim was to determine to what extent children in low-income schools learn novel, Tier 2 academic vocabulary words as a result of explicit vocabulary instruction, whether learning was consistent with hypothesized levels of knowledge, and whether higher levels of learning were moderated by prior vocabulary ability or special education status. Our specific research questions included the following:

1. At each unit test (four per year), did the vocabulary instruction group differ from a comparison group on vocabulary assessments reflecting different levels of word learning: word recognition (Level 2), receptive identification (Level 3), expressive labeling (Level 4), and decontextualized definitions (Level 4)?

We hypothesized better performance on measures representing Level 2 than Level 3 than Level 4 vocabulary knowledge and minimal word learning among children not receiving explicit vocabulary instruction.

2. Do growth models demonstrate changes in higher-level vocabulary learning over time in the two groups and is learning moderated based on initial language and cognitive measures or based upon special education status?

We hypothesized that children receiving special education would show poorer vocabulary learning. Given contradictions in prior research, we could not predict the moderating effects of initial vocabulary or language status.

Method

Participants

All first graders from two elementary schools (10 classrooms) in a Florida Panhandle school district were recruited to participate. The elementary schools provided educational services to children from low-income families and primarily served families who are eligible for reduced or free school lunch (over 90%). At the beginning of the year, 218 participants with consent were given the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007) and Expressive Vocabulary Test–Second Edition (EVT-2; Williams, 2007) for matching purposes. Because of the continued enrollment and departure of students within the schools, enrollment continued through the beginning of second grade, for a total of 282 eligible participants. The ethnic breakdown of the sample was 75% African American, 15% Caucasian, 5% Hispanic, and 5% mixed; 54% were male and 46% were female. Less than 0.5% were identified with limited English proficiency.

Participants included students with disabilities and students at high risk for developmental disabilities. Students with disabilities were defined as those with Individualized Education Plans (IEPs). Students with disabilities enrolled at least part time in general education classrooms were included. Records were kept on enrollment in special education

services, and IEPs were monitored to determine what related services were provided. A total of 38 participants (15.8%) had IEPs. There were 21 in the treatment group and 17 in the comparison group; 21 were classified as language impaired, seven were classified as speech impaired, three were classified as specific learning disabled, two were classified as developmental delayed, one was classified as autistic, one was classified as mentally handicapped, one was classified as emotionally handicapped, and two were classified as unspecified. No group differences were revealed in *t* tests for pretest measures. Records also tracked students who were retained in the same grade; children who repeated grades (total = 41) potentially were provided a second dose of the same treatment and thus were not included in the analyses.

In summary, there were 282 children with parent permission to participate in the study. With the exclusion of 41 students who repeated grades, results are based on a sample of 241 children. Mobility, a characteristic of most low-income schools, was high, and children who withdrew continued their participation if they re-enrolled in either of the two schools. Of the 241 participants, 46 (19%) received fewer than three waves of intervention (a half year or less). The 94 participants (39%) who received more than eight waves of intervention represent children who participated in all 3 years. The mean number of waves of intervention did not differ between groups, 7.15 (SD = 4.27) and 6.84 (SD = 4.11) for the vocabulary and comparison groups, respectively.

Experimental Design

First-grade students were randomly assigned to two treatment conditions that were administered simultaneously within classrooms: a vocabulary instruction group and a phonics instruction comparison group. The students were randomly assigned in first grade and then continued to receive the intervention in the same experimental condition in second and third grade. Students were tracked for up to 13 waves of data collection in first through third grade. Matching was used at the beginning of first grade to ensure stratification of samples based on baseline vocabulary ability (Shadish, Cook, & Campbell, 2002). Pretest PPVT-4 and EVT-2 standard scores were averaged for each student, and students within classrooms were rank-ordered based on their scores. Students were paired by proceeding down the ranked list, and the pair were then randomly assigned to groups based on the flip of a coin. During the first-grade year and the beginning of second grade, the children that were added were assigned to condition to equate the number of participants in condition within classrooms. Teachers and testers were blind to the students' experimental conditions. Table 1 summarizes pretest scores on standardized vocabulary, language, and cognitive measures. The groups did not differ in vocabulary and language assessments, but a significant difference in the Kaufman Brief Intelligence Test–Second Edition (KBIT-2; Kaufman & Kaufman, 2004) was revealed, with a larger effect size for nonverbal IQ (d = 0.40) than for verbal IQ (d = 0.30).

Procedure

The vocabulary instruction and the comparison condition were both administered simultaneously in "listening centers" that were largely independent of direct teacher involvement. Research assistants were responsible for running listening centers, whereas teachers mainly monitored the activity and only intervened if necessary. At their desks, students independently listened to prerecorded interactive activities and interventions through headphones for 20 min, the same activity was repeated 4 days per week. Each student was provided with a daily packet that contained a copy of a decodable book from the *Open Court Curriculum,* a worksheet, and a pencil. All students listened to the same decodable book and engaged in choral reading at the beginning of each session (approximately 5 min). Because each experimental condition had a CD player and wireless headphones to accommodate the students in that intervention, both conditions could be run concurrently. Students listened to the CD and completed the vocabulary or phonics worksheets independently following the instructions provided. A question-and-answer review session concluded each session.

Scripts for both conditions were developed according to the following guidelines. All scripts required active participation and responding from students by asking questions and providing time for spoken and written responses. Students completed worksheets developed in conjunction with the scripts. Books and worksheets used familiar pictures/icons as orienting cues, so when students were asked to follow along, they could check whether they were in the right place. All scripts contained a question-and-answer review session. All scripts presented information in the active verb voice to simplify syntactic demands. Eighteen weeks of instruction were presented in four 4- to 5-week units presented during the school year. Thus, there were four waves of intervention and data collection per year.

Vocabulary Instruction Condition

Choosing the Vocabulary Words

In Grades 1 and 2, the curriculum included one "anchor" (Tier 1) word from *Open Court* stories and six novel Tier 2 vocabulary words (two nouns, two verbs, and two adjectives). In Grade 3, the curriculum only included six Tier 2 words. Over 3 years, students were taught 324 challenging Tier 2 words. (Targeted words appear in Appendix A). The teachers were blind to the words that were part of our intervention. Because we selected Tier 2 words as our target words, it is unlikely that the teachers taught any of the words that were part of our intervention.

Tier 2 words occur frequently in print across a variety of literature and are contained in the oral and reading vocabularies of mature language users. Words at this level provide precision and nuance to oral language and written text. Thus, teaching Tier 2 vocabulary has the potential to have an impact on oral language as well as reading comprehension (Beck et al., 2002). The Academic Word List

Table 1. Baseline means and standard deviations for the experimental and comparison groups on vocabulary, language, and intelligence measures.

	Experimental group			Comparison group			
Measures	***n***	***M***	***SD***	***n***	***M***	***SD***	***t* test**
PPVT-4	117	89.7	12.9	113	87.9	11.1	1.11
EVT-2	117	91.2	14.2	113	89.4	13.2	0.96
CELF-4	95	91.3	16.2	103	89.2	15.0	0.92
KBIT-2-Composite	99	93.6	14.6	100	87.6	13.3	3.01**
KBIT-Verbal	99	92.3	13.9	100	88.4	12.1	2.11*
KBIT-Nonverbal	99	96.4	17.3	100	89.9	14.8	2.81**

Note. PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; EVT-2 = Expressive Vocabulary Test–Second Edition; CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition; KBIT-2 = Kaufman Brief Intelligence Test–Second Edition.

$^{**}p < .01$. $^{*}p < .05$.

(Coxhead, 2000) was developed from a list of the most frequently occurring words from introductory university texts across a variety of subjects. Words contained in this list fit the Beck et al. (2002) definition of Tier 2 words. There are approximately 570 root words, with the most frequently occurring variation of each word identified in the lists. By third grade, the researchers were running out of words that could be defined, illustrated, and fit into stories. Thus, the word selection was augmented using the Living Word Vocabulary List (Dale & O'Rourke, 1976). Stories were typically related to *Open Court* stories and were created with the constraint of using words selected from these lists that fit the story theme.

Creating Teaching Scripts

Every script contained multiple opportunities for engagement and response. Following the introduction of the definition of a word (anchor or targeted vocabulary word), children heard each word 12–15 times (four to six with associated definitions). They practiced saying the word (broken down if more than two syllables), repeating the definition, relating the word to the student's life, associating it with a gesture or a mnemonic phrase, and identifying the corresponding picture on the "activity sheet." (See Appendix B for an example of an instructional script.)

Phonics Instruction Comparison Condition

The comparison condition was presented in the same prerecorded format but provided instruction related to reading that was form-related rather than meaning-related. First-grade instruction focused on rhyming, alliteration, blending, and segmenting initially. Subsequent phonics instruction was adapted from *SPELL-Links to Reading and Writing* (Wasowicz, Apel, Masterson, & Whitney, 2004) and covered different orthographic rules/patterns, encoding patterns, and spelling patterns. Lessons progressed in difficulty from learning consonant sounds to explicit instruction on alternate spellings and syllabic spelling patterns.

Students listened to recorded scripts and completed corresponding worksheets by circling the correct answer. All scripts provided interactive opportunities for students to practice spelling, writing, and reading targeted encoding patterns. Students were given several opportunities to practice each of these skills during each session, for example:

> Say snake quietly to yourself. Snake starts with the /s/ sound. Look at the pictures next to the snake. Circle the picture that starts with the /s/ sound, like snake. Which picture did you circle? I hope you circled snail. Snail and snake both start with /s/. What sound do snake and snail start with? /s/, that's right!

Measures

Four curriculum-based assessments of targeted vocabulary words were used to monitor the acquisition of levels of word knowledge. These researcher-made measures mapped onto the levels of word learning outlined by Dale (1965). Although we present the measures below from lowest level of word learning to highest (i.e., Level 2 through Level 4), the four researcher-made measures were administered in the following order: recognition probes, decontextualized definition probes, expressive labeling probes, and last, the generalization task, which was the receptive identification probes. Thus, the order ensured that children did not receive an advantage on the decontextualized probes by being exposed to any pictures during the expressive or receptive probes, and the receptive task was administered last so the students did not receive any advantage on the expressive labeling probes by being reminded of the words during the receptive probes.

Recognition Probes

Recognition probes were administered to determine the student's ability to recognize the target word. The task required the student to identify the target word from two choices: the target word and a phonetically matched nonsense or rare word. For example, children were asked, "Which of these words have you heard before: demolish or puh-NEW-kuv?" This measure maps onto Level 2 of Dale's (1965) taxonomy: Has the child ever heard the word before, and if so, can he or she recognize it when presented

with a distractor word? A speech pathologist developed the foils with the criteria that the foil must have the same length and order of consonants and vowels.

Receptive Identification Probes

The receptive identification probes represented a generalization task to evaluate whether children had the depth of word knowledge to identify novel pictures of the words they learned. In other words, the receptive probes were not the same target pictures that children were exposed to during the intervention, so it was not a test of whether the child simply remembered the training pictures. The receptive probes mapped onto Level 3 of Dale's (1965) taxonomy. Receptive identification probes were modeled after the PPVT-4. Two training items familiarized participants with the test procedures and response requirements. Each testing item contained four illustrations: one that represented the target word and three additional pictures that served as foils. The foils included illustrations of a semantically related item, a phonologically similar item, and a conceptually unrelated item. An accurate response required the student to choose the correct picture depicting the trained vocabulary item. The position of the correct test object varied across the four positions on the page to minimize inference based on the placement of the items.

Target pictures and foils were developed using an iterative process. First, the measurement development team came up with a list of candidate foils to go along with each target word that were in the same word class. Pictures were selected from a slate of candidate items, and undergraduate and graduate research assistants were asked to select the target word and to explain what the answers they did not select represented. These procedures were followed to identify target pictures and to avoid foils that were potentially confusing because they were too similar to the target word. Efforts were made to select pictures that were line drawings as well as photographs. For a given word, all the choices were either line drawings or photographs. Once the slate of items for the unit was created, a final check ensured that illustrations of target words and distracters (i.e., foils) were transparent for young participants. The items were presented to two young children of the same age as the participants in the intervention to determine if they could select the target word. They were asked to explain why they selected their answer, and the child explained their process of elimination. Pictures that were not accurately identified as intended were replaced from a slate of distractor items that were created for the study.

Expressive Labeling Probes

Expressive labeling probes assessed students' ability to provide a label for the novel vocabulary word when given the trained picture stimulus with or without the definition. This measure maps onto Level 4 of Dale's (1965) taxonomy. The student viewed the picture of the target item that they were trained with during the intervention on a single plate and was asked to label the picture in response to the question, "What is this?" If the child did not provide a correct response, the child was asked, "What word means _____?" (e.g., What word means ten years?—for decade). A correct response required the student to provide the trained vocabulary word; all other responses were scored as incorrect.

Decontextualized Definition Probes

Decontextualized definition probes were administered to determine the student's level of word knowledge without contextual support from text or illustrations. These probes map onto Level 4 of Dale's (1965) taxonomy. The task required the student to explain the meaning of the novel vocabulary words (i.e., synonym or definitional knowledge) in response to the examiner's request to "Tell me everything you know about ___." If the child only provided one attribute of a word, the child was prompted further by saying, "Tell me something else about ____." The procedure and format of the decontextualized vocabulary probe is similar to that used in the Comprehensive Receptive and Expressive Vocabulary Test (Wallace & Hammill, 1994). A correct response required the student to provide at least one synonym or brief description (i.e., function, characteristic, etc.) of the target word; all other responses were scored as incorrect.

Unit Assessments

Decontextualized Definition and Expressive Labeling subtests for the words trained during that week were administered every Friday after 4 days of instruction to provide formative information. However, the primary dependent variables included in this report were unit assessments. Unit assessments included all four subtests and were administered every 5 weeks in the fall semester and every 4 weeks in the spring semester. The vocabulary words introduced during the unit were randomly ordered in each subtest. To minimize potential carryover learning effects among assessments, the order of administration was recognition, decontextualized, expressive, and receptive probes. Students did not receive any feedback on the accuracy of their responses.

Norm-Referenced Measures

Norm-referenced measures of vocabulary, language, and cognitive development were administered at the beginning of the study. The standardized measures of vocabulary include the PPVT-4 (Dunn & Dunn, 2007) and the EVT-2 (Williams, 2007). The Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wigg, & Secord, 2003) was used to assess oral language. The KBIT-2 (Kaufman & Kaufman, 2004) was administered to obtain an estimate of verbal and nonverbal intelligence.

Reliability

Reliability of measure instruments for the sample was estimated using alpha coefficients to estimate internal consistency. Internal consistency for decontextualized vocabulary probes were high and ranged from .82 to .89 for first grade, from .90 to .94 for second grade, and from .91 to .94 for third grade. Similarly, internal consistency for expressive vocabulary probes were high and ranged from .92 to .96 for

first grade, from .94 to .96 for second grade, and from .94 to .96 for third grade. Interrater reliability was calculated for 25% of sessions for measures requiring judgments of accuracy (e.g., decontextualized and expressive vocabulary probes). Two research assistants who were blind to the child's treatment group coded the data. Disagreements were resolved by one of the authors by reviewing the original data. We computed percentage agreement and kappa. As can be seen in Table 2, interrater agreement was high across grade levels.

Results

Group Differences on Four Measures of Vocabulary Learning

The first research investigated group differences on vocabulary assessments reflecting different levels of word learning at each assessment wave using *t* tests. Table 3 presents group means, standard deviations, sample size, and effect size results for the lower levels of vocabulary learning, the Word Recognition (top), and the Receptive Identification (bottom) measures. Chance responding is inherent in the Recognition subtest (two choices or 50% chance) and in the Receptive Identification subtest (four choices or 25% chance). In addition, in first and second grade, one of seven words (14%) was an anchor word. Thus, knowledge of novel words was expected to surpass 64% for the Recognition subtests and 39% for the Receptive subtest in first and second grade and 50% and 25%, respectively, in third grade. For the vocabulary group, mean Recognition subtest performance ranged from 78% to 98%, averaging 92% correct. For the comparison group, mean Recognition performance ranged from 61% to 82%, averaging 73% correct. Receptive subtest performance ranged from 59% to 86%, averaging 72% for the vocabulary group, versus from 40% to 61%, averaging 50% for the comparison group.

The *t* test results revealed that the children in the vocabulary instruction group performed significantly higher than the comparison group at each testing wave from Grade 1 through Grade 3. The vocabulary instruction group averaged a 19% advantage on the recognition subtest and a 22% advantage on the receptive subtest over the comparison group. Word recognition tended to improve after first grade, with a ceiling effect for the vocabulary group. Receptive identification also showed improvement after first grade. Effect sizes are averaged based on independent group *t* tests for the four unit tests per grade level. Effects sizes were quite large, ranging from 1.0 to 2.2 for the Recognition measure and from 0.6 to 2.6 for the Receptive measure.

Table 2. Interrater reliability for decontextualized definition and expressive vocabulary probes.

Reliability measures	Decontextualized probes	Expressive probes
First grade		
Percent agreement	1,363/1,677 = 81%	2,486/2,496 = 100%
Kappa	.83	1.00
Second grade		
Percent agreement	5,151/5,615 = 92%	5,366/5,394 = 99%
Kappa	.92	.99
Third grade		
Percent agreement	6,454/6,454 = 100%	6,442/6,442 = 100%
Kappa	1.00	1.00

Table 4 presents results for the Expressive Labeling and Decontextualized Definition measures. Independent groups *t* tests for each testing wave reflect large differences between experimental conditions. The vocabulary instruction group consistently out-performed the comparison group. The standard deviations are very large, however. For these comparisons, the homogeneity of variance assumption was rejected; thus, unequal variances *t* tests were conducted. All group differences were statistically significant, and effect sizes are presented at the bottom of Table 4. For the vocabulary group, mean Expressive Labeling subtest performance ranged from 21% to 57%, averaging 42%, and Decontextualized Definition subtest performance ranged from 12% to 37%, averaging 22%. As expected, the comparison group did poorly in labeling pictures and defining words they had not been taught, averaging 3% and 4%, respectively. Effects sizes were quite large, ranging from 1.4 to 3.1 for the Expressive Labeling measure and from 1.0 to 1.8 for the Decontextualized Definition measure.

Unit Growth Analysis for Expressive Labeling and Decontextualized Defining

Data Analysis Plan and Model Development

The second research question examined the longitudinal pattern of higher-level vocabulary learning in the two experimental conditions. As in the first research question, we expected much higher vocabulary performance in the group of children receiving explicit instruction. However, we sought to determine whether learning was moderated by standardized measures of vocabulary, language, and cognition at pretest and moderated by special education status.

To maximize the contributions of subjects with missing data in light of our high attrition rate over the 3 years of the study, hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) was used. These analyses compared groups for the two measures that represent higher levels of vocabulary knowledge: (a) the percentage correct for the Expressive Labeling measure and (b) the percent correct for "decontextualized, complete definitions." The overall question for each of the two vocabulary measures was: what trends were observed across the waves of data collection, and how could differences in these trends between the comparison group and the vocabulary instruction group be explained?

No actual baseline measure was given for the Expressive measure because all children during piloting scored a zero; thus, the baseline score was set to zero for all participants for this measure. The Decontextualized Definition

Table 3. Percent correct word recognition (top) and receptive identification (bottom) shown by unit and by intervention group.

Group		Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9	Wave 10	Wave 11	Wave 12	Average
							Word recognition measure							
Vocabulary	Mean	80%	78%	86%	88%	96%	96%	97%	98%	97%	97%	96%	96%	92%
	n	80	77	74	72	84	81	76	78	66	64	67	68	73.9
	SD	12.24	12.47	13.50	14.35	8.38	7.26	7.49	4.14	9.36	7.58	8.72	9.46	9.6
Comparison	Mean	61%	66%	63%	67%	72%	78%	73%	84%	78%	82%	74%	74%	73%
	n	71	79	76	83	83	80	75	76	53	51	47	46	68.3
	SD	12.42	10.60	9.76	15.74	12.93	10.62	14.27	10.80	14.76	14.00	13.96	13.62	12.8
	d	1.50	1.03	1.98	1.40	2.16	1.98	2.14	1.81	1.60	1.38	2.00	1.97	1.8
							Receptive identification measure							
Vocabulary	Mean	60%	59%	66%	60%	81%	86%	84%	80%	70%	73%	79%	68%	72%
	n	80	77	74	72	84	81	76	78	66	64	67	68	73.9
	SD	12.54	12.11	12.09	12.49	12.40	9.84	13.14	13.32	17.81	13.96	13.62	15.30	13.2
Comparison	Mean	44%	53%	53%	51%	53%	61%	54%	52%	42%	43%	54%	40%	50%
	n	71	79	79	83	83	80	75	76	53	51	47	46	68.6
	SD	11.31	10.12	15.69	12.32	8.38	11.19	10.07	12.39	10.17	12.24	14.37	10.12	11.5
	d	1.29	0.57	0.95	0.78	2.64	2.33	2.54	2.19	1.83	2.25	1.77	2.06	1.8

Note. Chance responding was 64% and 39% for recognition and receptive tasks for Waves 1–8 and 50% and 25% for Waves 9–12. Effect sizes (*d*) are reported for significant group differences based on *t* tests at each testing wave.

Table 4. Percent correct expressive labeling (top) and decontextualized definitions (bottom) shown by unit and by intervention group.

Group		Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9	Wave 10	Wave 11	Wave 12	Average
							Expressive labeling							
Vocabulary	Mean	23%	21%	27%	26%	54%	53%	57%	55%	48%	42%	49%	45%	42%
	n	64	78	75	73	83	80	75	78	66	64	67	68	72.6
	SD	20.30	17.21	22.26	21.17	24.66	21.26	23.24	24.80	24.52	20.75	21.75	26.68	22.4
Comparison	Mean	4%	1%	1%	2%	4%	9%	5%	3%	3%	3%	3%	2%	3%
	n	71	78	75	82	82	80	74	75	53	51	47	46	67.8
	SD	3.95	2.53	1.97	3.56	5.21	5.55	3.52	4.48	4.48	4.46	4.75	3.79	4.0
	d	1.36	1.62	1.67	1.65	2.82	2.88	3.10	2.91	2.47	2.47	2.70	2.07	2.3
							Decontextualized definitions							
Vocabulary	Mean	18%	12%	13%	12%	37%	26%	34%	32%	23%	16%	20%	16%	22%
	n	65	78	75	73	83	80	75	78	66	64	67	68	72.7
	SD	15.23	10.28	14.44	12.95	23.20	19.75	23.61	22.57	20.07	16.10	15.29	17.02	17.5
Comparison	Mean	4%	5%	3%	2%	8%	12%	4%	3%	2%	1%	1%	1%	4%
	n	71	78	75	82	82	80	74	75	53	51	47	46	67.8
	SD	4.94	3.79	3.55	4.68	8.80	8.54	3.86	3.86	2.03	1.16	1.87	1.58	4.1
	d	1.21	0.95	1.04	1.06	1.63	0.88	1.82	1.75	1.44	1.27	1.61	1.20	1.3

Note. Effect sizes (*d*) are reported for significant group differences ($p < .001$). Because homogeneity of variance assumption was rejected, unequal variances *t* tests were conducted.

measure included a Unit 1 baseline administration. To control for chance, anchor words were adjusted out (subtracted) prior to determining percent correct. After baseline, the unit tests were administered at 9, 16, 27, and 34 weeks (Year 1); 60, 66, 79, and 84 weeks (Year 2); and 113, 119, 132, and 139 weeks (Year 3).

The sample reported includes 241 children who did not repeat any grade during the course of the 3-year study. Baseline PPVT-4 and EVT-2 scores were taken as the first administration of these tests. Special education status (IEP) was obtained for each child in the sample (1 = *yes,* 0 = *no*). Other covariates included composite IQ based on the KBIT-2 and a composite language score based on the CELF-4.

Before applying covariates to subsequent analyses, simple comparisons of students with and without missing data on the KBIT-2 IQ measure and the CELF-4 Language measure were conducted to check for possible bias between intervention groups. Two covariates had excessive missing data, with approximately 17.8% (n = 43) of kids missing information on the KBIT-2 IQ or CELF-4 Language measures at baseline. In contrast, PPVT-4 and EVT-2 had only 4.6% (n = 11) missing data (and all students missing PPVT-4 were also missing EVT-2). All covariates were used as fixed predictors at the child level (i.e., these variables were not used as time-varying predictors). In HLM analyses, children with missing covariate data were excluded from Level 2 analyses. This suggests that the exclusion of the 50 children with missing scores on the KBIT-2 or CELF-4, or the PPVT-4 and EVT-2 variables may result in analyses that do not fully capture the relations between the predictors and the outcome of interest, despite the fact that the pattern of missing data was similar between the comparison and intervention groups. Thus, the inclusion of KBIT-2 and CELF-4, as well as PPVT-4 and EVT-2 measures, reduced the total number of children available for the growth models from 241 to 191.

Given the larger proportion of missing covariate data for the KBIT-2 and the CELF-4, we focused our missing data analyses on potential differences in study variables based on whether children were missing either of these two covariate scores. There was no association between the number of missing measures and treatment group (χ^2 = 2.096, p = .351). No significant differences in baseline PPVT-4 were found between children with zero (n = 191), one (n = 9), or two (n = 30) missing scores on either of the covariates, KBIT-2 or CELF-4 (p = .573). However, relative to children with no missing values on KBIT-2 or CELF-4, baseline EVT-2 scores were somewhat lower for children with both values missing on the KBIT-2 or the CELF-4 (p = .055). Of the four covariates, the only variable that showed initial differences between groups at baseline was the KBIT-2 (p = .003), with a mean IQ of 87.6 for the comparison group and a mean of 93.6 for the vocabulary instruction group. However, the effect size was small (η^2 = .044). Between students with and without IEPs, means for all four covariates were found to be significantly lower for students with IEPs (p < .002 for all).

Expressive Labeling and Decontextualized Definition showed slow increases in performance over time, followed by a decrease in third grade; this was more pronounced for the latter measure. The graphs of means suggested that a quadratic model would be the best way to fit a model to these higher-level measures of vocabulary. Consistent with the growth patterns evident in Figures 2 and 3, quadratic growth models were investigated. We began with a set of theoretically relevant predictors and used a parsimonious model-building approach advocated by Raudenbush and Bryk (2002) rather than a backward elimination approach where all potential Level 2 predictors are included at once. Before examining group differences, analyses were adjusted for baseline values of PPVT-4, EVT-2, CELF-4, and KBIT-2. Then, we retained significant covariates, and in subsequent models we examined all differences between groups and for whether children had an IEP.

For the final set of retained predictors, we examined interactions between IEP and intervention group, intervention group and PPVT-4, intervention group and EVT-2, and IEP interactions with PPVT-4 and with EVT-2.

HLM Results

The unconditional growth model and the final model that best fit the Expressive Labeling data are presented in Table 5 and depicted graphically in Figure 2. The unconditional growth model and the final model that best fit the Decontextualized Definition data are presented in Table 6 and depicted graphically in Figure 3.

As can be seen in Figure 2, having an IEP was not a significant predictor of the expressive vocabulary scores. However, there was a significant interaction effect. That is, children in the vocabulary instruction group who also had an IEP did not benefit as much as those without an IEP. Children in the comparison group showed no growth in expressive labeling of the novel words.

Figure 2. Graph of model for predicting expressive labeling showing effects of Individualized Education Plans (IEP) status.

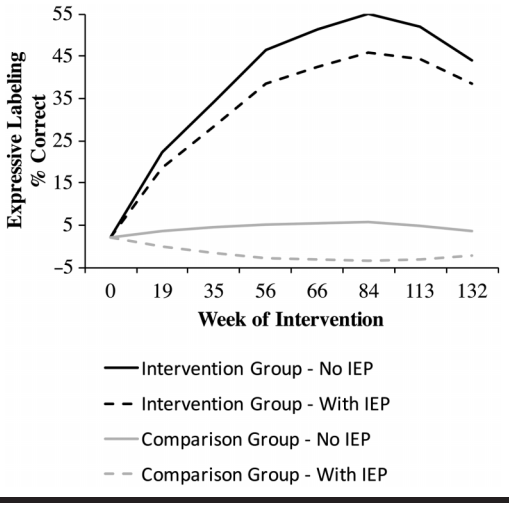
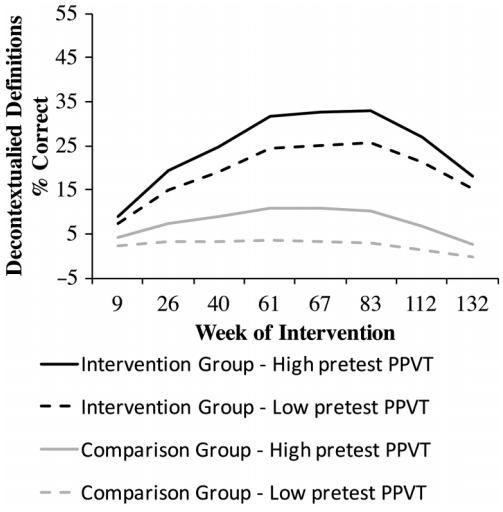


Figure 3. Graph of model for predicting decontextualized definitions showing effects of Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4) pretest scores.



Scores on the PPVT-4 was not a significant predictor of decontextualized definition scores. However, the interaction of PPVT-4 with group assignment was significant. As can be seen in Figure 3, children with higher PPVT-4 scores benefited more from vocabulary instruction than those with lower PPVT-4 scores.

Discussion

As expected, the two groups differed significantly on vocabulary assessments at each unit test for each measure reflecting deeper levels of knowledge, from word recognition to receptive identification to expressive labeling to decontextualized definitions. The effect sizes were large, with mean differences between groups averaging 19% for the word recognition tests and 21% for the receptive identification tests. The mean differences averaged 38% and 18% for the expressive labeling and decontextualized definition tests, respectively.

The inclusion of multiple measures gives an indication of the complexity of characterizing vocabulary learning. Vocabulary learning varied based on the level of vocabulary knowledge demonstrated by various measures. Of the 324 academic vocabulary words taught in 3 years, children in the experimental group recognized a mean of 298 words (92%). The large effect size (M = 1.75) reflects recognition of words 33% above chance compared to 13% above chance for the comparison group. The experimental group receptively identified a mean of 234 words (72%). The large effect size (M = 1.77) reflects receptive identification of words 38% above chance compared to 16% above chance for the comparison group. These group differences are not surprising, but we thought that the comparison group might have heard some of the target words before (word recognition), but that they would be less likely to demonstrate comprehension on the receptive identification task. However, when we take into account chance performance, the comparison group seemed

Table 5. Fixed effects maximum likelihood estimation for growth models showing the effects of IEP status on expressive labeling.

	Unconditional growth model		Final model	
Fixed effects[a]	**Coefficient**	***SE***	**Coefficient**	***SE***
Model for baseline (π_0)				
Intercept (β_{00})	2.42	0.59	2.36	0.60
Model for linear slopes (π_1)				
Intercept (β_{10})	0.56	0.04	0.07	0.02
KBIT-2 IQ (β_{11})			0.009	0.002
Intervention group (β_{12})			1.08	0.06
Special education status (β_{13})			0.001	0.03 *ns*
Intervention Group × Special Education Status (β_{14})			–0.37	0.14
Model for quadratic slopes (π_2)				
Intercept (β_{20})	–0.003	0.0003	–.0004 (.0001)	
KBIT-2 IQ (β_{21})			–.00004 (.00002)	
Intervention group (β_{22})			–.006 (.0004)	
Special education status (β_{23})			.00002 (.0002)	
Intervention Group × Special Education Status (β_{23})			.002 (.0009)	
Random effects (variance components)	**Variance**		**Variance**	
Variance in baseline (τ_{00})	(fixed to zero)		(fixed to zero)	
Variance in linear slopes (τ_{11})	0.44		0.11	
Variance in quadratic slopes (τ_{22})	0.00001		0.000001	
Variance within children (σ^2)	97.37		97.65	
Deviance (parameters)	13,928.41 (7)		13,684.45 (15)	

Note. IEP = Individualized Education Plan; KBIT-2 = Kaufman Brief Intelligence Test–Second Edition.

[a]All effects significant except where noted. Child-level predictors are grand mean–centered; categorical variables (special education status, intervention group) are dummy coded.

Table 6. Fixed effects maximum likelihood estimation for growth models showing the effects of baseline PPVT-4 scores on decontextualized definition performance.

	Unconditional growth model		Final model	
Fixed effects[a]	**Coefficient**	***SE***	**Coefficient**	***SE***
Model for baseline (π_0)				
Intercept (β_{00})	2.19	0.48	2.10	0.58
Model for linear slopes (π_1)				
Intercept (β_{10})	0.41	0.03	0.15	0.033
KBIT-2 IQ (β_{11})			0.002	0.001
Intervention group (β_{12})			0.52	0.04
PPVT-4 baseline (β_{13})			0.004	0.003 (p = .082)
Intervention Group × PPVT-4 (β_{14})			0.02	0.003
Model for quadratic slopes (π_2)				
Intercept (β_{20})	–0.003	0.0002	–0.001	0.0002
Intervention group (β_{21})			–0.003	0.0003
PPVT-4 baseline (β_{22})			–0.00004	0.00002
Intervention Group × PPVT-4 (β_{23})			–0.0001	0.00002
Random effects (variance components)	**Variance**		**Variance**	
Variance in baseline (τ_{00})	(fixed to zero)		(fixed to zero)	
Variance in linear slopes (τ_{11})	.18		.04	
Variance in quadratic slopes (τ_{22})	.00001		.000001	
Variance within children (σ^2)	90.23		90.17	
Deviance (parameters)	12,984.49 (7)		12,755.53 (14)	

Note. PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; KBIT-2 = Kaufman Brief Intelligence Test–Second Edition.

[a]All effects significant except where noted. Child-level predictors are grand mean-centered; categorical variables (intervention group) are dummy coded.

to perform similarly on these two tasks despite what we would consider different levels of knowledge, albeit an average of less than one word per week. Chance level of performance makes these results difficult to interpret with confidence. Nevertheless, there was clearly a ceiling effect for word recognition among second and third graders in the experimental group and consistently large group differences for word recognition and receptive identification tasks.

Chance performance is not a factor for the expressive labeling and decontextualized definition tasks. The experimental group expressively labeled a mean of 135 pictures (42%) and defined a mean of 70 words (22%). This compares to knowledge of a mean of 10–12 (3%–4%) words for the comparison group. Because of these multiple measures, these results provide an indication of deeper knowledge of learning of targeted vocabulary words. Performance of the various vocabulary assessments was consistent with our predictions with the best performance on the Level 2 recognition task and poorest performance on the Level 4 decontextualized definition task. Children who received vocabulary instruction demonstrated Level 2 and Level 3 knowledge of most words, even though Level 4 knowledge was not demonstrated for the majority of words. Effect sizes across 12 waves of data collection were large, averaging 1.8, 1.8, 2.3, and 1.3 for recognition, receptive identification, expressive labeling, and decontextualized defining tasks, respectively.

These longitudinal data demonstrated that growth was evident in academic vocabulary learning, but not exactly as predicted. The growth trajectory was best described by a quadratic function. As expected, improved learning was evident as children progressed from first to second grade. We would not have been surprised if this higher level of learning stabilized, but instead it declined in third grade. This pattern was most pronounced for the most demanding definition task, averaging 32% in third grade and 19% in third grade. For the expressive task, a smaller decrement in third grade was evident, stabilizing at 46% correct labeling after 55% in second grade.

The reason for the decrement is not clear. It is possible that the target vocabulary words were relatively more difficult in third grade. Indeed, as we began to exhaust words from the Academic Word List (Coxhead, 2000), we were left with more abstract words that were difficult to incorporate into stories and the Living Word List (Dale & O'Rourke, 1976) was used to augment word selection. Some support for this word difficulty hypothesis is evident when examining vocabulary performance for the comparison group. Although performing much lower than the vocabulary instruction group, the comparison group also performed better in second grade than third grade. Research is needed to determine how to scale words, specifically academic vocabulary words, for longitudinal intervention studies. The Coxhead (2000) approach of scaling words based on their frequencies in textbooks across disciplines seems like a reasonable approach to identifying potential target words. The vocabulary is naturally academic given the data source and the use across disciplines helps distinguish between what would be considered Tier 2 versus Tier 3 words. Of course, target word selection also must consider other factors, as interventionists must identify corresponding pictures (i.e., imageability) and definitions simple enough for children to understand.

A second possible reason for the decrement was the procedural change in third grade, as anchor words were no longer incorporated into training to accommodate longer stories that were used in the fluency practice for both groups. Although anchor words were excluded from vocabulary learning results, it could be that they serve a positive function. Ensuring a certain level of success with familiar words may enhance motivation in responding to more difficult, academic Tier 2 words.

In response to our second research question, growth models showed that both special education status (IEP) and initial receptive vocabulary performance (PPVT-4) affected vocabulary learning. Children in the vocabulary instruction group who also had an IEP did not benefit as much as those without an IEP. IEP status did not matter for children in the comparison group, as there was no growth in expressive labeling of the novel words. It is important to note, however, that children identified with special needs (as reflected by an IEP) still benefited from the vocabulary instruction, albeit not to the same extent as children developing typically. This study did not intervene to a predetermined criterion based on the outcome from the weekly assessments. All students, regardless of disability status, received new words and intervention scripts each week. It is possible that greater word learning would have resulted for students with defined disabilities if they received a more intensive, individualized intervention with increased exposures to the academic vocabulary (Foorman & Torgesen, 2001).

Likewise, children with higher PPVT-4 scores benefited more from vocabulary instruction than those with lower PPVT-4 scores. Thus, this study helps to clarify the moderating role of initial vocabulary status. Results were consistent with studies that also have found that children with lower vocabulary scores made smaller gains in learning (Coyne et al., 2007; Penno et al., 2002; Pullen et al., 2010). It could be that discrepant results in the literature are due to differences in the difficulty of the targeted vocabulary or perhaps the measures used to assess learning. One might speculate that little difference might be seen in learning highly imageable, concrete vocabulary words (typically nouns), and more differences might be seen in the learning of more abstract, less imageable words. For the latter, foundational vocabulary is likely to be more important. For the former, there may be more room for growth among children who have had less opportunity to learn from a linguistically rich home language environment.

Because few studies have implemented and evaluated long-term, targeted vocabulary interventions, it is difficult to compare the effects of this intervention with other similar interventions. It is not surprising to find that explicit instruction is superior to no explicit instruction on vocabulary words, as large effects have been reported consistently (e.g., Coyne et al., 2007; Zipoli et al., 2011). This study is unique in several respects. This study tracked vocabulary learning longitudinally over 3 years. Furthermore, this report focuses on the number of words learned rather than simply reporting group differences. Because vocabulary intervention studies typically include estimates of effect sizes and rarely report findings in terms of the number or percentage of words learned, it is difficult to determine whether this vocabulary intervention was especially effective. Comparing across studies is complex. Differences in the words taught, the population of interest, vocabulary measures used, the context and dosage of instruction, as well as the intervention strategies employed all may affect results. Kelley et al. (2015) estimated that first graders receiving vocabulary intervention learned an average of 27% of 10 vocabulary words taught by Penno et al. (2002). Supplemental instruction on eight vocabulary words provided by Loftus et al. (2010) resulted in an average of 28% correct on an expressive definition measure. Kelley et al. (2015) also point out that students may experience considerable training before demonstrating these seemingly low learning rates. In our study, vocabulary training comprised about 1 hr per week for 18 weeks and a total of 108 words per year. Despite the large scope of this study in terms of the number and selection differences in the words taught and relatively brief instruction on a per word basis, vocabulary learning results appear comparable to previous research. 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.

The academic vocabulary words selected for this intervention were challenging. Despite being characterized as Tier 2 words (Beck et al., 2002), these words rarely seemed to be encountered in kindergarten to third grade enough to be learned by children in the comparison group; they learned to label 10 of the targeted words and to define 12 words in 3 years, whereas the intervention group learned to label 135 words and to define 70 words on average. Although this word learning accounted for a small portion of the vocabulary that students were expected to learn during the year (i.e., 3,000 words according to Nagy, 2005), its compounding effects on academic achievement may have long-lasting benefits. Future investigations should evaluate the effects that longitudinal vocabulary interventions that teach Tier 2, academic vocabulary words, have on reading comprehension and academic achievement in the later school years for children at risk for language and literacy difficulties.

Assessment of levels of word knowledge is an ongoing challenge for researchers and educators attempting to determine the depth of vocabulary knowledge gained. The NRP (2000) has suggested that researcher-made measures are needed for lack of standardized measures that are likely to be sensitive to intervention effects. Given the nature of vocabulary learning, there is little reason to suspect that vocabulary learning will have much effect on vocabulary tests such as the PPVT-4 and the EVT-2, unless words taught are included in the norm referenced test. Although the vocabulary instruction group improved by an average of 5.1 and 7.4 standard score points on the PPVT-4 and EVT-2, respectively, by the end of the study, there was no significant difference between treatment groups. These differences also may reflect that the remaining sample of participants came from more stable home environments, given the high attrition

rates, rather than intervention effects. It may be more important to determine whether vocabulary instruction ultimately benefits later academic achievement. We have a preliminary indication of potential long-term benefits as the vocabulary instruction group significantly outperformed the comparison group in Rasch units at the end of third grade ($d = 0.42$) on the Measures of Academic Progress, a computer-adaptive assessment of reading achievement (NWEA, 2009). Given the high attrition rate and lack of corroborating evidence, however, this finding should be considered speculative.

When it comes to relatively high-level academic words, exposure alone does not appear to result in much learning. Therefore, the focus of future vocabulary instruction and intervention efforts should consider trying to maximize the results of explicit instruction. Future research is needed to inform what words should be taught, how many words should be taught, how to best assess acquisition, and how to maximize learning. For example, one apparent shortcoming of the instructional procedure employed was the lack of review over time to lessen the chances of children forgetting or confusing previously taught words. More information is needed on the extent to which words are retained. This is important if vocabulary instruction is expected to have long-term effects on reading fluency and later academic achievement as hypothesized.

The results of this study should be interpreted with regard to a number of limitations. Researchers and graduate students managed this automated intervention. This limits generalizability of the findings to classroom teachers or other school staff. Future research should evaluate implementation of this intervention by classroom teachers and paraprofessionals to determine if the intervention would have similar effects.

Unfortunately, we were unable to follow participants to determine the impact of the participants' word learning on their later reading comprehension. Although the attrition rate was rather typical of high-poverty schools, the cumulative effects of attrition are indicative of the difficulty in gathering longitudinal data with this population; 48% of our original sample remained after 3 years.

Conclusion

Producing substantial gains in language and vocabulary growth in children from poverty is critical to prevention and early intervention efforts and is likely to require a long-term approach. Clearly, efficient interventions are needed to overcome these significant differences in developmental trajectories. This study provides an opportunity to analyze word learning at the item level to inform the process of revising the curriculum to make it more efficient. Future research may also consider more individualized approaches to developing higher levels of learning.

In summary, this study exemplifies several laudable characteristics. By evaluating the effects of vocabulary instruction over 3 years, we could describe vocabulary learning over time. In particular, we were able to describe absolute word learning (i.e., average number of words learned) on multiple measures. This also allowed for an examination of the depth of word learning, and results clearly indicated different levels of word learning on measures that tapped lower versus higher levels of word knowledge. Despite considerable attrition, a rigorous experimental design allowed results to be presented in growth curve analyses that also provided an examination of moderating effects. Differential responses to intervention were revealed based on both entry-level vocabulary knowledge as well as IEP status.

The current study adds to the emerging evidence on explicit vocabulary instruction for children in the early elementary grades. The method of direct, explicit teaching of academic vocabulary embedded in children's stories seems to show great promise for students at risk for language or literacy difficulties. Indeed, the use of recorded lessons and minimal instructional materials seems feasible for widespread application with children in high-poverty schools. Our explicit vocabulary instruction provided ample opportunities for children to respond to instruction individually, albeit at the same time. It is critical that academic vocabulary be an instructional focus for students at risk for reading delays or academic underachievement. On the basis of the results of this study, multiple practice opportunities coupled with carefully constructed instructional scripts have the potential to help reduce the gaps in vocabulary development evident across socioeconomic class in the early primary grades.

Acknowledgments

This research was partially supported by a grant from the U.S. Department of Education, Institute of Education Sciences (R324L060023) awarded to Florida State University. We gratefully acknowledge the cooperation and assistance of the staff, students, and parents in Leon County Schools. We greatly appreciate the assistance of Ann O'Connell with statistical analyses.

References

Anderson, R. C., & Nagy, W. E. (1991). Word meanings. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 690–724). Mahwah, NJ: Lawrence Erlbaum.

Apthorp, H. S. (2006). Effects of a supplemental vocabulary program in third-grade reading/language arts. *The Journal of Educational Research, 100*(2), 67–79.

Bailey, A. L., & Heritage, M. (2008). *Formative assessment for literacy, grades K-6: Building reading and academic language skills across the curriculum* (pp. 61–91). Thousand Oaks, CA: Corwin Press.

Baker, S. K., Simmons, D. C., & Kame'enui, E. J. (1998). Vocabulary acquisition: Research bases. In D. C. Simmons & E. J. Kame'enui (Eds.), *What reading research tells us about children with diverse learning needs* (pp. 183–218). Mahwah, NJ: Lawrence Erlbaum.

Baumann, J. F., & Graves, M. F. (2010). What is academic vocabulary? *Journal of Adolescent and Adult Literacy, 54,* 4–12.

Baumann, J. F., & Kame'enui, E. J. (Eds.). (2004). *Vocabulary instruction: Research to practice*. New York, NY: Guilford.

Beck, I. L., & McKeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 789–814). White Plains, NY: Longman.

Beck, I. L., & McKeown, M. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107,* 251–272. https://doi.org/10.1086/511706

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

Beck, I. L., McKeown, M. G., & Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 147–163). Hillsdale, NJ: Erlbaum.

Biemiller, A. (2004). Teaching vocabulary in the primary grades: Vocabulary instruction needed. In J. F. Baumann & E. J. Kame'enui (Eds.), *Vocabulary instruction: Research to practice* (pp. 28–40). New York, NY: Guilford.

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade instruction. In A. Hiebert & M. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 223–242). Mahwah, NJ: Erlbaum.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98,* 44–62. https://doi.org/10.1037/0022-0663.98.1.44

Bolger, D., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance based learning approach. *Discourse Processes, 45*(2), 122–159. https://doi.org/10.1080/01638530701792826

Bortfeld, H., Morgan, J., Golinkoff, R., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science, 16,* 298–304. https://doi.org/10.1111/j.0956-7976.2005.01531.x

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49,* 278–293. https://doi.org/10.1044/1092-4388(2006/023)

Cain, K., & Oakhill, J. (2011). Matthew-effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities, 44,* 431–443.

Christ, T. (2011). Moving past "right" or "wrong" toward a continuum of young children's semantic knowledge. *Journal of Literacy Research, 43*(2), 130–158. https://doi.org/10.1177/0022219411410042X11403267

Cirrin, F., & Gillam, R. (2008). Language intervention practices for school-aged children with spoken-language disorders: A systematic review. *Language, Speech, and Hearing Services in Schools, 39,* 110–137. https://doi.org/10.1044/0161-1461(2008/012)

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

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

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

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12*(3), 145–162. https://doi.org/10.1207/s15327035ex1203_3

Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English, 42,* 895–901, 948.

Dale, E., & O'Rourke, J. (1976). *The living word vocabulary: The words we know, a national vocabulary inventory*. Elgin, IL: Dome.

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development, 82,* 870–886. https://doi.org/10.1111/j.1467-8624.2011.01576.x

Duff, F. J., & Hulme, C. (2012). The role of children's phonological and semantic knowledge in learning to read words. *Scientific Studies of Reading, 16,* 504–525. https://doi.org/10.1037/e512592013-035

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56*(8), 848–856. https://doi.org/10.1111/jcpp.12378

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

Ehri, L. C. (1998). Grapheme–phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Erlbaum.

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

Fenson, L., Dale, P., Reznick, J. S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs for the Society for Research in Child Development, 59*(5, Serial No. 242). https://doi.org/10.2307/1166093

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all students. *Learning Disabilities Research and Practice, 16*(4), 203–212. https://doi.org/10.1111/0938-8982.00020

Graves, M. F. (2016). *The vocabulary book: Learning and instruction* (2nd ed.). New York, NY: Teacher's College Press.

Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review, 31*(3), 328–349.

Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., Golinkoff, R. M., & Nesbitt, K. (2015). Examining the acquisition of vocabulary knowledge depth among preschool students. *Reading Research Quarterly, 51*(2), 181–198.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Brookes.

Hatch, E., & Brown, C. 1995. *Vocabulary, semantics, and language education*. New York, NY: Cambridge University Press.

Hiebert, E. H., & Lubliner, S. (2008). The nature, learning, and instruction of general academic vocabulary. In A. E. Farstrup & S. J. Samuels (Eds), *What research has to say about vocabulary instruction* (pp. 106–129). Newark, DE: International Reading Association.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74,* 1386–1378. https://doi.org/10.1111/1467-8624.00612

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. https://doi.org/10.1044/0161-1461(2005/003)

Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of*

Speech, Language, and Hearing Research, 53, 739–756. https://doi.org/10.1044/1092-4388(2009/08-0248)

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test–Second Edition*. Bloomington, MN: Pearson.

Kelley, E. S., & Goldstein, H. (2014). Building a Tier 2 intervention: A glimpse behind the data. *Journal of Early Intervention, 36*(4), 292–312. https://doi.org/10.1177/1053815115581657

Kelley, E. S., Goldstein, H., Spencer, T., & Sherman, A. (2015). Effects of automated Tier 2 storybook intervention on vocabulary and comprehension learning in preschool children with limited oral language skills. *Early Childhood Research Quarterly, 31,* 47–61. https://doi.org/10.1016/j.ecresq.2014.12.004

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32,* 69–92. https://doi.org/10.1017/s0142716410000299

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

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

Marzano, R. J. (2003). Direct vocabulary instruction: An idea whose time has come. In B. Williams (Ed.), *Closing the achievement gap* (pp. 48–56). Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., & Pickering, D. J. (2005). *Building academic vocabulary: Teacher's manual*. Alexandria, VA: Association for Supervision and Curriculum Development.

Maynard, K. L., Pullen, P. C., & Coyne, M. D. (2010). Teaching vocabulary to first-grade students through repeated shared storybook reading: A comparison of rich and basic instruction to incidental exposure. *Literacy Research and Instruction, 49,* 209–242.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47,* 91–108. https://doi.org/10.1002/rrq.011

Nagy, W. E. (2005). Why vocabulary instruction needs to be long-term and comprehensive. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 27–44). Mahwah, NJ: Erlbaum.

Nash, H., & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language and Communication Disorders, 41,* 335–354.

Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research, 48,* 439–458. https://doi.org/10.1044/1092-4388(2005/030)

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Jessup, MD: National Institute for Literacy.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards*. Washington, DC: Author.

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

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*(1–2, Serial No. 149). https://doi.org/10.2307/1165788

Neuman, S. B. (2013). Giving all children a good start: The effects of an embedded multimedia intervention for narrowing the vocabulary gap before kindergarten. *Technology as a Support for Literacy Achievements for Children at Risk, Literacy Studies, 7,* 21–32. https://doi.org/10.1007/978-94-007-5119-4_3

Neuman, S. B., & Wright, T. S. (2013). *All about words: Increasing vocabulary in the common core classroom, PreK-2*. New York, NY: Teachers College Press.

NWEA. (2009). *Technical manual for Measures of Academic Progress for primary grades*. Lake Oswego, OR: Northwest Evaluation Association.

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*(4), 763–782. https://doi.org/10.1111/1467-8624.00498-i1

Payne, A. C., Whitehurst, G. W., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly, 9,* 427–440. https://doi.org/10.1016/0885-2006(94)90018-3

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

Phythian-Sence, C., & Wagner, R. K. (2007). Vocabulary acquisition: A primer. In R. Wagner, A. Muse, & K. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 1–14). New York, NY: Guilford.

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58,* 510–519. https://doi.org/10.1598/rt.58.6.2

Puhalla, E. M. (2011). Enhancing the vocabulary knowledge of first-grade children with supplemental booster instruction. *Remedial and Special Education, 32*(6), 471–481.

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 and Practice, 25*(3), 110–123. https://doi.org/10.1111/j.1540-5826.2010.00309.x

Raikes, H., Pan, B. A., Luze, G., Tamis-LeMonda, C. S., Brooks-Gunn, J., Constantine, J., . . . Rodriguez, E. T. (2006). Mother–child book reading in low-income families: Correlates and outcomes during the first three years of life. *Child Development, 77,* 924–953. https://doi.org/10.1111/j.1467-8624.2006.00911.x

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (advanced quantitative techniques in the social sciences)*. Thousand Oakes, CA: Sage.

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

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development, 83,* 508–525. https://doi.org/10.1111/j.1467-8624.2011.01710.x

Saffran, J. R., Werker, J., & Werner, L. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In R. Siegler & D. Kuhn (Eds.), *Handbook of child development* (pp. 58–108). New York, NY: Wiley.

Scarborough, H. S. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothesis with some ugly facts. In H. W. Catts & A. G. Kamhi (Eds.), *The*

connections between language and reading disabilities (pp. 3–24). Mahwah, NJ: Erlbaum.

Scott, J. A., Jamieson-Noel, D., & Asselin, M. (2003). Vocabulary instruction throughout the day in twenty-three Canadian upper-elementary classrooms. *The Elementary School Journal, 103,* 269–286. https://doi.org/10.1086/499726

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals Preschool–Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Sénéchal, M., & Cornell, E. H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly, 28,* 360–374. https://doi.org/10.2307/747933

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Silverman, R. (2007). Vocabulary development of English language and English only learners in kindergarten. *The Elementary School Journal, 107*(4), 365–383. https://doi.org/10.1086/516669

Snow, C. E. (1994). Beginning from baby talk: Twenty years of research on input and interaction. In C. Gallaway & B. Richards (Eds.), *Input and interaction in language acquisition* (pp. 3–12). United Kingdom: Cambridge University Press.

Spencer, E. J., Goldstein, H., Sherman, A., Noe, S., Tabbah, R., Ziolkowski, R., & Schneider, N. (2012). Effects of an automated vocabulary and comprehension intervention: An early efficacy study. *Journal of Early Intervention, 34*(4), 195–221. https://doi.org/10.1177/1053815112471990

Stahl, S. A. (1986). Three principles of effective vocabulary instruction. *Journal of Reading, 29,* 662–668.

Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings*. Mahwah, NJ: Erlbaum.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407. https://doi.org/10.1598/rrq.21.4.1

Steele, S. C., & Watkins, R. V. (2010). Learning word meanings during reading by children with language learning disability and typically-developing peers. *Clinical Linguistics and Phonetics, 24,* 520–539. https://doi.org/10.3109/02699200903532474

Swanson, H. L. (1986). Do semantic memory deficiencies underlie learning readers' encoding processes? *Journal of Experimental Child Psychology, 41,* 461–488. https://doi.org/10.1016/0022-0965(86)90005-6

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Walker, D., Greenwood, C., Hart, B. & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development, 65,* 606–621. https://doi.org/10.2307/1131404

Wallace, G., & Hammill, D. D. (1994). *Comprehensive Receptive and Expressive Vocabulary Test*. Austin, TX: Pro Ed.

Wasowicz, J., Apel, K., Masterson, J., & Whitney, A. (2004). *SPELL-Links to reading and writing*. Evanston, IL: Learning by Design.

Williams, K. (2007). *Expressive Vocabulary Test–Second Edition*. Bloomington, MN: Pearson.

Wright, T. S., & Neuman, S. B. (2014). Paucity and disparity in kindergarten oral vocabulary instruction. *Journal of Literacy Research, 46*(3), 330–357. https://doi.org/10.1177/1086296x14551474

Zipoli, R. P., Coyne, M. D., & McCoach, D. B. (2011). Enhancing vocabulary intervention for kindergarten students: Strategic integration of systematically related an embedded word review. *Remedial and Special Education, 32*(2), 131–143. https://doi.org/10.1177/0741932510361262

Appendix A (p. 1 of 2)

Vocabulary Words Targeted in Each Unit

Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6
accessible	adjust	adjacent	alternative	advertisement	abandoned
adequate	affordable	airplane[a]	city[a]	advise	accumulate
amiss	conflict	assistant	collapse	aloof	attach
baby[a]	coordinate	camera[a]	conclusion	ancient	attempt
bond	crab[a]	component	demonstrate	appearance	berries[a]
challenging	distribute	display	duration	appreciative	bird[a]
convert	domestic	fee	eager	astonished	camouflage
crouch	earn	finalize	evidence	audience	castle[a]
crucial	export	focus	exposed	citizen	coward
design	external	function	feature	concentrate	deceive
devices	gender	gradual	flag[a]	corporation	defined
devote	global	illustrate	kite[a]	decade	desire
differentiate	grandmother[a]	immense	location	decline	detect
diverse	guideline	indicate	mend	dinner	elated
egg[a]	initiate	isolated	negative	disturb	error
emerge	job[a]	library[a]	odd	encounter	examine
facilitate	manual	map[a]	option	excessive	fond
food[a]	modify	persuade	prohibited	exclude	frightened
fragile	objective	proceed	react	fortune	generous
individual	occupy	relocate	release	home[a]	horizon
inquire	overseas	research	reluctant	journey	invisible
inspect	practitioner	respond	rely	mature	irate
involve	previous	route	require	mysterious	lizard
meadow	professional	section	spider[a]	notion	observe
migrate	pursue	significant	symbol	offer	parcel
outcome	resources	sufficient	technique	parallel	permanent
phase	sector	visual	unified	party[a]	pleasant
predator	specific	voluntary	utilize	perplexed	receipt
rabbit[a]	tradition			prefer	schedule
residence	transit			rain[a]	scold
similar	travel[a]			rare	slope
tranquil	ultimate			relative	toy[a]
weary	undertake			seek	tremble
web	vehicle			shoes[a]	wicked
welfare	worker			substitute	wound

(*table continues*)

Appendix A (p. 2 of 2)

Vocabulary Words Targeted in Each Unit

Unit 7	Unit 8	Unit 9[b]	Unit 10[b]	Unit 11[b]	Unit 12[b]
bedroom[a]	assume	analyze	accurate	abrupt	apprehension
boast	assure	anticipate	anticipation	altercation	bizarre
century	brief	approximately	consistent	appeal	collaborate
commence	bundle	aware	contract	barter	conspicuous
companion	canal	commitment	contribution	catastrophe	deluge
constant	caution	confirm	convene	compensate	ecstatic
correspond	complex	debate	convincing	console	expedition
cottage	compliment	discreet	cooperative	dejected	illusion
dinosaur[a]	considerate	economical	determined	demolish	inquisitive
elongate	day[a]	emphasis	devoted	escalate	jeopardy
friends[a]	deviate	enforce	equilibrium	essential	luminous
game[a]	distinct	illegal	establish	feat	orient
infer	domineering	immigrants	expertise	gratuity	plead
intervene	dragon[a]	income	fluctuate	indecisive	prevail
laborious	emphasize	justify	highlight	memorable	proximity
modern	expandable	lecture	ignorant	persevere	replica
numerous	field[a]	liberate	inhibited	profitable	ruminate
ordinary	indicator	majority	legislator	quest	serene
peculiar	inform	minority	propose	remorse	shriek
persistent	interaction	motivation	publish	resolution	succumb
region	loyal	neutral	reassure	restrain	unwillingness
rival	manual	nuisance	regulation	strive	utter
scheme	miserable	persist	reveal	superb	vast
scientist	principle	resolve	revise	zany	wary
source	quarrel	swiftly	secure		
suffice	reject	task	shift		
variable	signify	tender	terminate		
wander	teacher[a]	tense	undertaking		
		transmit	version		
		unfortunate	vigorous		

[a]Anchor words. [b]Anchor words were not used in these units.

Appendix B

Vocabulary Instruction Script and Instructional Sequence

An *Open Court* story was always read first for decoding practice in Grade 1 and for reading fluency practice in Grades 2 and 3. The scripts and anchor words were linked to the *Open Court* story theme (e.g., the anchor word was "airplane" in "The Plane Trip" script).

> NARRATOR: Today Ben is going to take us on an *airplane* ride. Use your pencil to find the picture of the *airplane*. If your pencil is on the right spot, make your hand move like a plane flying in the sky.

Each script contained a regular sequence of script elements for each word, all incorporating interactive activities. Transitions from one vocabulary word to the next were accomplished using the script story.

> **Script introduction:** NARRATOR: *Hi there! Welcome to the listening center. We have lots of things to do today. Do you have today's reading book? Do you have a pencil? You're going to need your work papers too. Get ready! Here we go!*
>
> **Fluency practice:** NARRATOR: *Pick up your reading book. The name of this book is, "The Bug, the Duck, and the Frog." Put your finger on the first word. Let's read the first word. The first word is "All." Let's keep reading. Remember to turn the pages when you hear the gong sound.*
>
> **Word introduction:** CHARACTER: *Hi guys! My name is Mandy and soon it will be my tenth birthday! I can't wait to have my birthday **party**, except for one thing. My mom says I have to invite the new girl in my class, and I don't want to.*
>
> **Direction to picture and verbal/moto-kinesthetic action:** NARRATOR: *Put your pencil on the people in the birthday hats. If your pencil is on **party,** say, "It's party time!!"*
>
> **Word repetition, relate to picture, and word repetition:** NARRATOR: *I want you to say **party** quietly. Circle the cake and say the new word again. It's **party**—you're amazing!*
>
> **Definitions and context:** NARRATOR*: A **party** is a fun get-together with friends. People have all sorts of **parties**. Have you ever been to a birthday **party?** Cool! A birthday **party** is a fun get-together with friends where you eat yummy cake and play games. Next weekend, Mandy's going to have a fun get-together for her birthday. She is going to invite her friends to her **party**—but her mom says she has to invite the new girl in her class, too!*
>
> **Expressive naming task:** CHARACTER: *Next weekend, I'm going to have a… a… what do you call a fun get-together with your friends? A **party**—you are such a big help! Tell me again what a **party** is. A fun get-together with your friends—right on!*
>
> **Transition to next word:** CHARACTER: *My mom said I have to invite the new girl to my **party,** but I don't want her to come. She never talks to anyone and I don't think she's very nice. So, I want to **exclude** her from my party! When you leave someone out, you **exclude** them.*
>
> **Word review:** Words were reviewed in a different order than in the script. Words that started with the same sound(s) were separated to decrease confusion. The review included reference to the worksheet picture (e.g., Put your finger on "task"), reference to moto-kinesthetic/verbal "fun thing," and decontextualized and expressive tasks.