Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: www.tandfonline.com/journals/uree20

Direct and Extended Vocabulary Instruction in Kindergarten: Investigating Transfer Effects

Michael D. Coyne, D. Betsy McCoach, Susan Loftus, Richard Zipoli Jr., Maureen Ruby, Yvel C. Crevecoeur & Sharon Kapp

To cite this article: Michael D. Coyne, D. Betsy McCoach, Susan Loftus, Richard Zipoli Jr., Maureen Ruby, Yvel C. Crevecoeur & Sharon Kapp (2010) Direct and Extended Vocabulary Instruction in Kindergarten: Investigating Transfer Effects, Journal of Research on Educational Effectiveness, 3:2, 93-120, DOI: 10.1080/19345741003592410

To link to this article: https://doi.org/10.1080/19345741003592410

Journal of Research on Educational Effectiveness, 3: 93–120, 2010

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345741003592410

Routledge
Taylor & Francis Group

INTERVENTION, EVALUATION, AND POLICY STUDIES

Direct and Extended Vocabulary Instruction in Kindergarten: Investigating Transfer Effects

Michael D. Coyne and D. Betsy McCoach
University of Connecticut, Storrs, Connecticut, USA

Susan Loftus
University of Rhode Island, Kingston, Rhode Island, USA

Richard Zipoli Jr.
West Hartford Public Schools, West Hartford, Connecticut, USA

Maureen Ruby
Eastern Connecticut State University, Willimantic, Connecticut, USA

Yvel C. Crevecoeur
The City College of New York, New York, New York, USA

Sharon Kapp
Marlborough Public Schools, Marlborough, Massachusetts, USA

Abstract: The purpose of this study was to investigate the efficacy of an 18-week program of direct and extended vocabulary instruction with kindergarten students on both proximal measures of target word knowledge and transfer measures of generalized language and literacy. A second purpose was to examine whether treatment effects would be moderated by initial receptive vocabulary knowledge measured at pretest. In a quasi-experimental design, 80 kindergarten students from schools serving large at-risk populations were taught the meanings of 54 vocabulary words within interactive story read alouds over 36 half-hour instructional lessons (2 lessons per week over

Address correspondence to Michael D. Coyne, Department of Educational Psychology, University of Connecticut, 249 Glenbrook Road, Unit 2064, Storrs, CT 06269, USA. E-mail: mike.coyne@uconn.edu

18 weeks). An additional 44 students served as a no-treatment control. Findings indicated that students who received vocabulary instruction outperformed controls on a measure of target word knowledge as well as measures of generalized receptive vocabulary and listening comprehension. In addition, initial receptive vocabulary was strongly related to posttest performance on all measures. Implications are discussed in relation to supporting vocabulary development in the early grades within a multitier framework of instruction and intervention.

Keywords: Vocabulary instruction, kindergarten students, listening comprehension, storybook reading

The relationship between vocabulary knowledge, reading comprehension, and academic achievement is well established (Becker, 1977; National Reading Panel, 2000). Recent longitudinal research, in particular, has underscored the importance of early vocabulary development to future reading success. For example, converging evidence suggests that although vocabulary knowledge develops relatively independently from beginning reading skills such as decoding in the early grades, it becomes increasingly predictive of reading comprehension over time (Catts, Hogan, & Adolf, 2005; Share & Leiken, 2004; Storch & Whitehurst, 2002; Vellutino, Tunmer, Jaccard, & Chen, 2007). Moreover, vocabulary knowledge in the early grades continues to predict reading comprehension in third grade and beyond (Cunningham & Stanovich, 1997).

In response to growing evidence highlighting the importance of early vocabulary development, researchers have called for more teacher directed vocabulary instruction, especially in the primary grades (Biemiller, 2001). There have been a number of recent intervention studies evaluating vocabulary instruction in kindergarten and first grade (Beck & McKeown, 2007; Biemiller & Boote, 2006; Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009; Justice, Meier, & Walpole, 2005; Silverman, 2007). These studies have provided strong and converging evidence that young school-age students can learn the meanings of sophisticated vocabulary through direct instruction. However, these studies were designed to evaluate the impact of vocabulary instruction primarily on measures of words targeted for instruction and did not include transfer measures of generalized language or literacy. In addition, many of these studies found that direct vocabulary instruction was differentially effective favoring students with higher initial levels of vocabulary knowledge (Coyne et al., 2009; Penno, Wilkinson, & Moore, 2002). The purpose of this quasi-experimental study was to evaluate the impact of an 18-week program of direct vocabulary instruction in kindergarten on both proximal measures of target word knowledge as well as transfer measures of overall vocabulary knowledge and listening comprehension. A second purpose was to examine whether treatment effects would be moderated by initial receptive vocabulary knowledge.

DIRECT VOCABULARY INSTRUCTION WITH PRIMARY-GRADE STUDENTS

Numerous recent studies have evaluated the efficacy of direct vocabulary instruction on the target word learning of young school-aged students. This growing body of research has begun to identify components of effective vocabulary instruction for young children. For example, research has produced converging evidence that young children can learn the meanings of novel vocabulary when they are provided with student friendly definitions or simpler synonyms of target words within the context of a storybook or a meaningful sentence (Elley, 1989; Justice et al., 2005; Penno et al., 2002). In addition, research suggests that word learning can be enhanced when students are provided with extended opportunities to interact with target words in varied contexts beyond those from the original story (Beck & McKeown, 2007; Biemiller & Boote, 2006). In our own research with kindergarten students, we have found that an additional investment in instructional time that allows for increased encounters and exposures to target words leads to a greater depth of word knowledge (Coyne et al., 2009; Coyne, McCoach, & Kapp, 2007; see also Beck, McKeown, & Kucan, 2002).

In general, vocabulary intervention research with younger students converges with findings from research with older students and suggests that instruction that includes these components results in reliable and consistent increases in knowledge of words targeted for instruction (National Reading Panel, 2000; Stahl & Fairbanks, 1986). However, unlike vocabulary research with older students, few studies with younger children have included transfer measures of generalized vocabulary knowledge or comprehension. For example, in a recent meta-analysis, Elleman, Lindo, Morphy, and Compton (2009) identified 37 intervention studies that measured the impact of vocabulary instruction on measures of reading or listening comprehension. Of these 37 studies, only 3 included students in first grade or younger.

Evidence of transfer effects is important because of inherent limitations associated with direct vocabulary instruction. Although there is an ongoing discussion about the rate of vocabulary development, it appears that children in the primary grades add at least 2.2 new root words, on average, to their vocabulary in a day (Biemiller & Slonim, 2001), and this number may be as high as 9 (Carey, 1978). Because vocabulary instruction is time intensive, only a small percentage of the words students learn over the course of a year can be introduced directly through instruction (Anderson & Nagy, 1992). A much stronger case could be made for the worth of vocabulary instruction in the early grades if there was evidence to suggest that there were benefits beyond just learning the meanings of words targeted for instruction. Therefore, to fully evaluate the overall impact of direct vocabulary instruction with primary grade students, it is important for research to examine not only target word learning but also evidence of transfer effects on generalized language and literacy outcomes.

TRANSFER EFFECTS OF VOCABULARY INSTRUCTION

There are several mechanisms that could explain why directly teaching target vocabulary could theoretically produce transfer effects on more distal measures of language and literacy. First, vocabulary instruction can theoretically improve both listening and reading comprehension. The instrumentalist hypothesis is perhaps the most straightforward explanation for this relationship (Anderson & Freebody, 1981; Stahl, 1991). If students don't know the meanings of individual words, it interferes with their ability to understand the overall meaning of a sentence or paragraph. If meanings of many of the words in a passage are unknown or a single unknown word is particularly important to the overall meaning of a sentence, comprehension may not occur at all. Therefore, directly teaching the meanings of words that appear in a story or passage should improve students' comprehension of that passage. Even if students gain only partial knowledge of words, it may be enough to aid comprehension depending on the importance of the word and the supportiveness of the context (Stahl, 1991). Knowing the meanings of more of the words in a story or passage may also increase the likelihood that a student will be able to better use context to infer the meanings of other unknown words (Shefelbine, 1990).

Direct vocabulary instruction could also increase general word consciousness or metalinguistic awareness. Metalingustic awareness is the ability to reflect on and manipulate various features of language in an active and considerate manner (Nagy, 2007). Extended vocabulary instruction is characterized by teacher-supported activities that draw attention to words and their meanings (Beck et al., 2002). This instruction allows students to interact with and process words in different contexts, relate them to and discriminate them from other words, and connect them to their own prior knowledge. It is possible that this type of deep and extended vocabulary instruction could facilitate the development of metalinguistic awareness, especially in young students who may not have systematically attended to words and their meanings before. When students develop metalinguistic awareness and begin to consider words more carefully, they may become more independent and active word learners. Increased metalinguistic awareness may increase the likelihood that students will infer the meanings of unknown words in non-teacher-directed learning opportunities. Students with greater metalinguistic awareness may also become better able to consolidate, integrate, access, or verbally demonstrate the word knowledge they already possess (Van Kleeck, Stahl, & Bauer, 2003).

It is also possible that learning target words helps students refine and consolidate their knowledge of other, related words. In connectionist models of vocabulary development, learning new word meanings helps refine the lexical constraints of other partially known words that are semantically or situationally associated with the newly learned words (Landauer & Dumais, 1997). Because learners assume that no two words can mean exactly the same thing, new word learning helps to define and constrain the semantic space of all other related

words in the learner's lexicon (Lockett & Shore, 2003; Shore & Kempe, 1999). Therefore, by acquiring deep knowledge of words targeted for instruction, students may be able to consolidate their knowledge of other connected words, even if those words have not been encountered during instruction.

There is some evidence that vocabulary instruction with older students can result in gains in comprehension and general vocabulary knowledge. In their seminal review of vocabulary intervention studies, Stahl and Fairbanks (1986) reported an average effect size of 0.97 for researcher developed comprehension measures and 0.30 for global measures of comprehension. In a recent meta-analysis, however, Elleman et al. (2009) used a more conservative methodology and found a mean effect of 0.50 for custom comprehension measures that was statistically significant ($p < .01$) and an effect of only 0.10 for standardized measures of comprehension that was not statistically significant. Elleman et al. also found a statistically significant effect of 0.29 on measures of general vocabulary knowledge.

Few studies with younger students, however, have attempted to measure transfer effects, primarily because of the short duration of those studies. Elleman et al. (2009) found only three vocabulary intervention studies conducted with students in first grade and younger that included measures of comprehension. The findings of these studies were mixed. For example, according to calculations reported in Elleman et al., of the two studies that included custom comprehension measures, one found a statistically significant effect size of .59 with first grade students (Brabham & Lynch-Brown, 2002) and the other found a nonstatistically significant effect of .02 with preschool students (Kimer-Simon, 2003). The third study (Puhalla, 2005) used a standardized measure of comprehension with first grade students and found a non-statistically significant and negative effect of –.26.

DIFFERENTIAL EFFECTIVENESS OF VOCABULARY INSTRUCTION

Research on shared storybook reading suggests that students who are at risk for language and literacy difficulties with lower initial vocabularies are less likely than their peers with higher vocabularies to learn words incidentally while listening to stories (Coyne, Simmons, Kame'enui, & Stoomiller, 2004; Nicholson & Whyte, 1992; Robbins & Ehri, 1994; Senechal, Thomas, & Monker, 1995). There is also some evidence that the treatment effects of direct vocabulary instruction with young students may be moderated by individual differences in language and literacy measured at pretest (Coyne et al., 2009; Coyne et al., 2007; Penno et al., 2002), although a few studies have reported divergent findings (Coyne et al., 2004; Elley, 1989).

PURPOSE OF THE STUDY

In this study, we were interested in investigating the efficacy of an 18-week program of direct and extended vocabulary instruction with kindergarten students not only on proximal measures of target word knowledge but also transfer measures of generalized language and literacy. To do this, we included posttest measures of overall receptive vocabulary knowledge (i.e., Peabody Picture Vocabulary Test [PPVT-III]; Dunn & Dunn, 1997) and listening comprehension (i.e., an adapted version of the Strong Narrative Assessment Procedure [SNAP], that included words targeted for instruction; Strong, 1998). We also included an experimenter-developed measure of metalinguistic awareness or word consciousness that assessed students' ability to infer the meanings of novel vocabulary within the context of supportive sentences. We were also interested in whether students most at risk responded differentially to extended vocabulary instruction on both target word measures and transfer measures. Overall receptive vocabulary knowledge is a strong predictor of language and comprehension outcomes (Scarborough, 2005; Storch & Whitehurst, 2002) and responsiveness to vocabulary intervention (Coyne et al., 2009; Coyne et al., 2007; Penno et al., 2002). Therefore, we used students' PPVT scores assessed at pretest as an indicator of risk.

METHOD

Participants

Participants in this study were kindergarten students who attended three elementary schools in three different school districts in the Northeast. Each school serves large percentages of students who would be considered at risk for experiencing literacy and language difficulties based on demographic data and performance on state reading assessments. Schools A and B were located in small urban districts, and School C was located in a large urban district. School demographic data are presented in Table 1.

Table 1. School demographics

		Ethnicity					
School	Enrollment	Hispanic	African American	Caucasian	Other	Free or Reduced Lunch	% at Goal on State Mastery Test[a]
School A	291 (K-4)	60%	8%	29%	3%	81%	20%
School B	276 (K-5)	41%	38%	17%	4%	91%	36%
School C	552 (K-8)	71%	23%	5%	1%	73%	22%

[a]State average = 58%.

Permission forms were sent home to all kindergarten students in the three schools. A total of 134 permission forms were returned, which represented of a return rate of more than 85%. In consultation with teachers, 6 students were not included in the study based on significant behavioral and/or intellectual disabilities. Over the course of the study, 4 students moved (2 treatment, 2 control), resulting in 124 students completing the study. Eighty students were in the treatment group (40 male, 40 female), and 44 students were in the control group (23 male, 21 female). In the treatment group, 55% of students were Hispanic ($n = 44$), 21% were Caucasian ($n = 17$), 19% were African American ($n = 15$), and 5% were of other ethnicity ($n = 4$). In the control group, 57% of students were Hispanic ($n = 25$), 16% were Caucasian ($n = 7$), 25% were African American ($n = 11$), and 2% were of other ethnicity ($n = 1$). In the treatment group, 31 students (39%) were English learners and 17 students (39%) in the control group were English learners. Students in all schools received instruction only in English.

Design

To address our research questions, we used a quasi-experimental design. At Schools A and B, intact kindergarten classrooms were assigned to either treatment ($n = 56$ students) or control conditions ($n = 22$ students). At School A, two classrooms were assigned to implement the experimental vocabulary intervention and one classroom was assigned to serve as a no-treatment control. At School B, one classroom implemented the vocabulary intervention and one classroom served as a control. Interventionists were the classroom teachers and delivered the instruction to all students during whole class sessions. The three treatment teachers in Schools A and B had an average of 19 years of teaching experience, with 14 years of experience teaching kindergarten students. The two control teachers had an average of 26 years of teaching experience, with 13 years of experience teaching kindergarten students. All classroom teachers had master's degrees.

At School C, because of different scheduling constraints, all students across the three kindergarten classrooms were randomly assigned at the student level to either receive the vocabulary intervention ($n = 23$) or serve as a control ($n = 21$). Students in the treatment group received the intervention in small groups of 3 or 4 outside of the classroom. Interventionists were two graduate student-interns in the final year of an integrated bachelor's–master's degree program in special education. Each interventionist taught three intervention groups. Both interns had completed their student-teaching experience. Classroom teachers at School C had an average of 17 years of teaching experience, with 10 years of experience teaching kindergarten students.

Description of the Intervention

We developed our intervention, which we describe as extended vocabulary instruction, to provide students with direct instruction of target vocabulary words that is interactive, robust, and varied. An extended approach is characterized by instruction that introduces students to target vocabulary within the supportive context of a storybook but also provides extended opportunities to discuss and interact with target words in multiple and novel contexts outside of the story (see also Beck et al., 2002). Extended instruction encourages deep processing of word meanings that challenges students to move beyond memorizing simple dictionary definitions to understanding words at a richer, more complex level. We have evaluated the effects of this instruction on the short-term word learning of target vocabulary in previous studies (Coyne et al., 2009; Coyne et al., 2007).

The intervention consisted of 36 half-hour instructional lessons (2 lessons per week over 18 weeks). We developed the intervention around 18 storybooks that were read aloud to students followed by interactive postreading activities. The story-reading portion of the intervention averaged 10 to 20 min per session, and the postreading activities averaged 10 to 15 min. Each storybook was read twice during a week. The identical intervention was implemented in Schools A, B, and C. We selected storybooks that would be of high interest to kindergarten students and that included engaging narratives, rich language, and illustrative pictures. Priority was given to storybooks that depicted multicultural characters and themes. We selected three target vocabulary words from each storybook to teach directly (54 total). Our goal was to identify target words that were unfamiliar to students but whose meanings they would be able to understand. These are the features of words that Beck et al. (2002) characterized as Tier II words, words they recommend for teaching directly to primary-grade students. To accurately assess the strength of the intervention, we chose words that typical kindergarten students would be unlikely to know according to the *Living Word Vocabulary* (Dale & O'Rourke, 1981). The 54 target words included 15 nouns, 18 verbs, and 21 adjectives. We adapted storybooks minimally to control the number of exposures to each target word (i.e., once in each storybook). We also ensured that target words were close to a supportive illustration. On a few occasions, we also substituted lower frequency synonyms for story words. A list of storybooks and target vocabulary words are included in the appendix.

At the beginning of each storybook reading session, interventionists introduced the three target words and had students pronounce them. Students were then encouraged to listen for each of the *magic words* during the story reading and to raise their hands whenever they heard one. Interventionists acknowledged students when they recognized target words and then reread the sentence containing the word (e.g., "Oh, good. Some of you raised your hands! What word did you hear? Yes, *peculiar*. 'When Anansi said the word seven, a peculiar thing happened.'"). Interventionists then provided students with a simple definition of the word (e.g., "*Peculiar* means *strange or different*.") Next, interventionists reread the sentence and replaced the target word with

its definition (e.g., "Now I'll say the sentence again with the words that mean *peculiar*. 'When Anansi said the word seven, a *strange* thing happened.'"). Interventionists then reinforced the story context by referring to the illustration, "In the picture you can see that something strange or peculiar happened to Anansi that sent him flying through the air." Finally, students again pronounced the target word to strengthen phonological representations (e.g., "Everyone say *peculiar*."). The purpose of these procedures was to offer students both a simple definition of each target word as well as contextual support for the word's use in the story (Stahl & Fairbanks, 1986). The story reading portion of the intervention averaged 10 to 20 min per session.

After each storyreading, interventionists engaged students in activities that provided them with opportunities to interact with target words in rich and varied contexts beyond those offered in the story (Beck et al., 2002). First, students were reintroduced to each target word by reviewing how it was used in the story (e.g., "One of the magic words we learned in the story was *peculiar*. *Peculiar* means strange or different. 'When Anansi said the word seven, a peculiar thing happened.' In the picture you can see that something strange or peculiar happened to Anansi that sent him flying through the air."). Next, interventionists provided students with examples of the target word used in other contexts (e.g., "Other things could be peculiar too. If a person were wearing a bathing suit outside on a snowy day, they would look peculiar. If you saw a cat driving a car, that would be a peculiar thing. If a boy always walked backwards, that would look peculiar or strange and different.").

Students then engaged in different interactive activities that focused on recognizing examples and nonexamples of target words used in different contexts (e.g., "I'll show you some pictures. If you think the picture shows something that looks peculiar, or strange and different, put your thumbs up like this and whisper, 'That looks peculiar.' If the picture doesn't show something that looks peculiar, put your thumbs down like this and don't say anything."). Interventionists reinforcing correct responses (e.g., "If you put your thumb up like this, you're right! The kids in this picture look peculiar, or strange and different. 'Aren't Silly Sally and her friends peculiar? They're walking upside down on their hands.'") and followed up with open-ended questions that encouraged students to extend and elaborate on their initial responses (e.g., "Why do you think walking upside down is peculiar?"). Open-ended questions were designed to elicit an extended response that demonstrated a full understanding of the target word (e.g., "People don't walk upside down! That would be really strange."). Finally, interventionists provided additional individual turns to two or three students who may have been having difficulty. The post reading vocabulary activities averaged 10 to 15 min per session.

The vocabulary intervention also included systematic and distributed review of target words. The 54 target words were assigned to one of two review conditions that varied slightly in the type of review or a no-review comparison. Each of the 36 words in the two review conditions were practiced in approximately five additional lessons beyond the ones in which they were

initially introduced. Similar to initial word instruction, review activities reinforced definitions and encouraged deep processing of word meanings through opportunities to interact with words in novel contexts and explore the relationships between target words and previously known words and concepts. (A detailed comparison of the different review conditions can be found in Zipoli, Coyne, & McCoach, in press. In summary, students learned target words that received systematic review to a greater extent than words that received no review with effect sizes ranging from .69 to 1.03.)

Fidelity of Implementation

Prior to the start of the study, interventionists (i.e., kindergarten classroom teachers at Schools A and B, and graduate student interventionists at School C) participated in a 1-day training. During the training, members of the research team modeled vocabulary lesson activities and provided interventionists with opportunities for supervised practice with immediate corrective feedback.

To document and assess treatment fidelity, we developed an observation checklist that included critical components of the vocabulary intervention. (Gersten et al., 2005; Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). Examples of critical components incorporated into the fidelity checklist included whether or not interventionists delivered each instructional element, modeled procedures appropriately, maximized opportunities to respond, provided error correction, and read storybooks with enthusiasm. Interventionists were observed by project staff five times, on average, over the course of the study. Interobserver agreement for the fidelity observations was high (coefficient kappa-n of .97 for the story readings and .94 for the extension activities). Fidelity of implementation averaged above 95% for each interventionist indicating a high degree of treatment integrity. Researchers met with interventionists after each observation and provided feedback on implementation.

Based on observations, we were also able to estimate the average amount of instructional time spent on each target word over the course of the intervention. Each of the 36 words that were reviewed systematically received between 13 and 20 min of instruction and the 18 target words that were not reviewed received approximately 11 min of instruction.

We also observed storybook-reading activities in the two control classrooms at Schools A and B and the three kindergarten classrooms at School C using the same observation checklist. Our goal was to evaluate program differentiation (Dane & Schneider, 1998), or whether these teachers incorporated any elements of the direct vocabulary instruction into their typical story-reading activities. We observed in each classroom twice over the course of the study. Observation data indicated that storybook-reading activities in these classrooms focused primarily on story elements. We did not document any direct vocabulary instruction during our observations of storybook readings in control classrooms.

Measures

Our goal was to evaluate the effects of extended vocabulary instruction with both proximal measures that were closely aligned with the intervention as well as distal measures that would assess transfer to other areas of language and literacy. To assess the proximal effects of the intervention, we developed a measure of target word knowledge. To assess transfer, we measured general vocabulary knowledge, listening comprehension, and students' ability to infer the meaning s of unknown words in context.

Measure of target word knowledge. The measure of target word knowledge assessed students' understanding of vocabulary words taught directly during the intervention. We selected a sample of 37 of the 54 target words to include on this measure to reduce the length of administration. We sampled words that were representative of all word types and that were introduced throughout the 18-week intervention. Students were first asked to provide a definition of a target word (e.g., "Tell me what the word *halt* means."). Next, students were asked a question in which the target word was used in a relatively neutral context (e.g., "What would you be doing if you were *halting*?"). Each question was scored with 2 points for a complete response (e.g., "Halt means to stop"), 1 point for a partial or related response (e.g., "Halt is like a statue"), and zero points for an unrelated response, or no response. The maximum score a student could receive for each word was 4 (i.e., 2 points possible for each question). The internal consistency reliability for this measure was .98. We administered the definition question for each word at pretest to document any prior knowledge of target words and to examine comparability across treatment and control groups. We administered the both questions at posttest to evaluate impact of the intervention on target word knowledge.

Measure of general vocabulary knowledge. We used the PPVT–III (Dunn & Dunn, 1997) to assess students' generalized vocabulary knowledge. The PPVT is a norm-referenced measure of receptive vocabulary. Students are presented with four pictures and are asked to point to the picture that best represents the word given by the examiner. Standardized scores ($M = 100$, $SD = 15$) are computed based on number of items correct and the student's chronological age. Alternate-forms reliability coefficients for the PPVT range from .88 to .96, and test–retest reliability coefficients range from .91 to .94. Overall, research had suggested high reliability and validity for the PPVT (Salvia & Ysseldyke, 1998). We administered the PPVT at pretest to characterize students' overall receptive vocabulary knowledge prior to the intervention. We also administered an alternate form of the PPVT at posttest to assess the impact of the intervention on general vocabulary knowledge. Different forms of the PPVT were used at pretest and posttest. None of the target vocabulary words appeared on the PPVT.

Measure of listening comprehension. We used an adapted version of the SNAP (Strong, 1998) to assess students' listening comprehension. In the SNAP, students listen to a story read aloud while following along with pictures in a wordless storybook. We adapted the SNAP by incorporating 18 of our target words into the story, which had 395 total words. We selected target words that were representative of the words taught throughout the intervention and could be integrated in the story. Each of the words that were incorporated into the SNAP was also assessed as part of the target word measure. Following the reading, students were asked 10 questions about the story that were both literal and inferential. We adapted 5 of the questions so that they included a target word in the question (e.g., What happened when Joe *grasped* at the frog?) Answers to questions were scored as either correct or incorrect. The internal consistency reliability for this measure was .69. Although this is clearly below desirable levels, Gersten et al. (2005) asserted that reliabilities higher than.60 can be considered acceptable for newly created measures used in intervention research and indicate that a coherent measurement construct is being measured. We assessed listening comprehension only at posttest.

Measure of metalinguistic awareness. We developed an experimental measure to assess students' ability to infer the meanings of unknown words in context. Students first listened to a sentence containing a word unknown by most kindergarten students while looking at a picture that illustrated the sentence (e.g., "The bag full of water plunged to the ground and made a big splash"). Students were then asked to provide a definition of the unknown word (e.g., What do you think the word *plunged* means in that sentence?) Definitions were scored with 2 points for a complete response, 1 point for a partial or related response, and zero points for an unrelated response, or no response. There were eight items on the measure. The internal consistency reliability for this measure was .65, again, lower than ideal, but higher than .6 (Gersten et al., 2005). We assessed metalinguistic awareness only at posttest.

Data Collection and Scoring

Data collection took place at pretest and posttest. Members of the research team administered all assessments individually to students in quiet locations. We collected pretest data approximately 1 week prior to the start of the intervention. We collected posttest data within 1 week after the conclusion of the intervention. We collected pretest assessments in two 20-min session and posttest assessments in three 20-min sessions. All assessments were conducted in English.

Data collectors were required to demonstrate at least 90% reliability for administration. All measures were scored by one member of the research team. The project director randomly selected and independently scored 20% of the assessment protocols to check for scoring reliability. Agreement was 100%.

RESULTS

Analysis Plan

We initially included delivery method (teacher-led whole-class intervention in Schools A and B and the graduate student led small-group intervention in School C) and the delivery method by Treatment interaction into all of our models. None of these effects were statistically significant. In addition, we ran separate correlations between pretest PPVT scores and posttest measures for each delivery method (whole class, small group) to test whether there were different Treatment × Aptitude interactions across conditions. None of the correlations were statistically significantly different from each other. Because the intervention was standardized and implemented with similar procedures across all groups and there was no evidence of different treatment effects across delivery method, we subsequently analyzed the teacher-led whole-class intervention in Schools A and B and the graduate student led small-group intervention in School C together to increase the power of our design.

To determine the impact of the vocabulary intervention, we estimated the effect of the treatment after controlling for initial PPVT score. Originally, we had intended to use an analysis of covariance to analyze our data. However, we failed to meet the assumption of homogeneity of regression slopes across treatment and control groups. Instead, we consistently found an interaction between the covariate (initial PPVT) and treatment. Therefore, we used multiple regression to analyze our data. For each dependent variable, we estimated a model that included the treatment condition, the centered initial PPVT score, and the Treatment × PPVT interaction.

Because students were nested within classes and the intervention was delivered by teachers, we expected that the data were nonindependent. However, because we had only eight interventionists participating in the study, we could not run a multilevel analysis. In their simulation study, Maas and Hox (2005) found that multilevel models with 10 Level 2 units produced biased standard error estimates for the regression coefficients and variance components, as well as Level 2 variance components that were overestimated by approximately 25%. Instead, we computed the intraclass correlations (ICCs) for each of the dependent variables that we used in our analyses. The ICCs were 0 for the fall PPVT scores, fall target word measure scores, and spring listening comprehension task, .089 for the spring PPVT scores, and .289 for the spring target word measure scores. We then applied corrections to the *t* tests from the multiple regression analyses to take into account the nonindependence of data and to protect against an inflated Type I error rate. We used the correction procedure suggested by the What Works Clearinghouse (2008). In addition, we computed the design effect for each variable and weighted the standard errors in our models by the design effects. The ICCs were computed by teacher using the unconditional random effects models. Both procedures gave us very similar

Table 2. Pretest scores

	Control		Treatment	
	M	*SD*	*M*	*SD*
Target word measure (definition question)	4.83	3.87	5.35	6.05
Peabody Picture Vocabulary Test pretest	89.66	14.97	93.91	17.16

results. In Tables 5 through 8, we report the results of the original *t* tests and the corrected *t* tests computed using the formulas from the What Works documentation. We set our *p* value at .05. In addition, because of the small sample size and the exploratory nature of the study, we interpret effects that are not statistically significant at the .05 level but have an effect size above 0.25 as "substantively important" in accordance with What Works Clearinghouse (2008) procedures (p. 22).[1]

Results

We administered the target word measure and the PPVT prior to the start of intervention to document any prior knowledge of target words and to examine comparability across treatment and control groups. There were little differences between the treatment and control on the target word measure, $t(125) = .52$, $p = .60$, or on the Standardized PPVT at pretest, $t(123) = 1.38$, $p = .17$, indicating comparability between groups. The descriptive statistics for the pretest scores are shown in Table 2. The ICCs for the two pretest variables were 0; therefore, no correction was applied to the *t* tests for the pretest measures.

At posttest we assessed target word knowledge, overall vocabulary knowledge (PPVT), listening comprehension, and metalinguistic awareness (i.e., students' ability to infer the meanings of unknown words). Descriptive statistics for the posttest scores are shown in Table 3. In addition, Table 4 contains the correlations among all measures.

The results of these regression analyses are contained in Tables 5, 6, 7, and 8. In addition, we computed Cohen's *d* effect sizes to examine the magnitude

[1]Using the What Works Clearinghouse documentation, the correction for clustering for the *t* test is

$$t_A = t\sqrt{\frac{(N-2)-2\left(\frac{N}{m}-1\right)\rho}{(N-2)\left[1+\left(\frac{N}{m}-1\right)\rho\right]}}$$

The correction to the degrees of freedom associated with the *t* tests corrected for clustering is

$$h = \frac{\left[(N-2)-2\left(\frac{N}{m}-1\right)\rho\right]^2}{(N-2)(1-\rho)^2+\frac{N}{m}\left(N-2\frac{N}{m}\right)\rho^2+2\left(N-2\frac{N}{m}\right)\rho(1-\rho)}$$

Table 3. Posttest scores

	Control		Treatment	
	M	*SD*	*M*	*SD*
Target word measure	9.70	5.45	55.50	37.58
Peabody Picture Vocabulary Test posttest	92.51	10.43	98.99	13.68
Listening comprehension	2.48	1.53	3.32	2.58
Metalinguistic awareness	5.40	3.40	6.86	3.95

Table 4. Correlations among the dependent variables

	1	2	3	4	5	6
1. PPVT pretest	1.0					
2. Target word pretest	.44	1.0				
3. PPVT posttest	.75	.46	1.0			
4. Target word posttest	.57	.36	.64	1.0		
5. Metalinguistic awareness	.51	.30	.43	.59	1.0	
6. Listening comprehension	.58	.31	.56	.64	.49	1.0

Note. All correlations are statistically significant at $p < .01$. PPVT = Peabody Picture Vocabulary Test.

Table 5. Regression analysis: Target word measure coefficients

		Unstandardized		Standardized			Corrected	
Model		B	SE	β	*t*	Sig.	*t*	Sig.
1	(Constant)	39.992	2.831		14.126	<.001		
	PPVT pre centered	1.334	0.179	.565	7.472	<.001		
2	(Constant)	10.059	3.490		2.882	.005		
	PPVT pre centered	0.157	0.256	.067	.614	.541	0.26	.80
	Treatment	43.770	4.278	.553	10.231	<.001	4.38	<.01
	Interaction	1.407	0.295	.515	4.772	<.001	2.04	.04

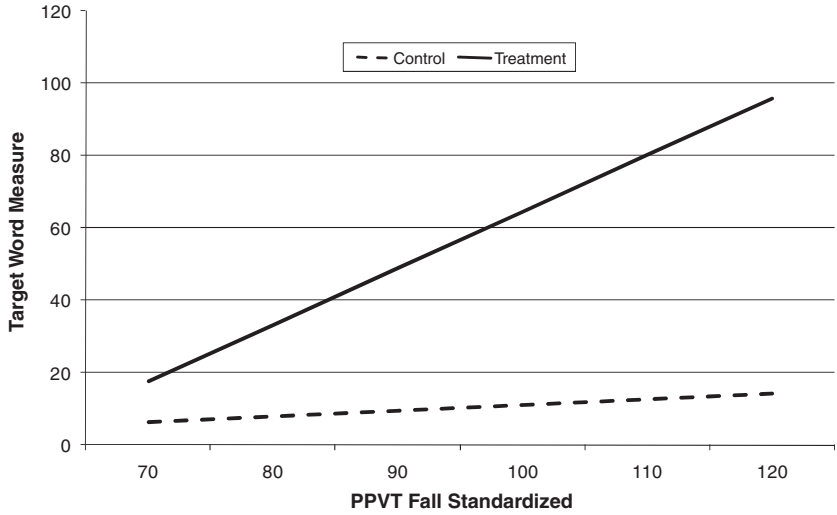

Figure 1. Regression analysis for the target word measure. *Note*. PPVT = Peabody Picture Vocabulary Test.

of the treatment effects. However, because of the interaction between treatment and initial PPVT score, the predicted magnitude of effect differed as a function of initial PPVT score. Therefore, we also computed predicted Cohen's *d* effect sizes at three different pretest PPVT scores: 85, 100, and 115. These effect sizes are contained in Table 9.

Table 5 reports the results of our regression analysis for the target word measure. These results are also displayed in Figure 1. Overall the treatment group outperformed the control group (corrected $p < .01$), and there was an interaction between pretest PPVT scores and the treatment (corrected $p = .04$). In other words, there were large differences between the treatment group and the control group on our measure of target word learning, and these differences

Table 6. Regression analysis: PPVT posttest coefficients

		Unstandardized		Standardized			Corrected	
Model		B	SE	β	t	Sig.	t	Sig.
1	(Constant)	96.933	0.793		122.301	<.001		
	PPVT pre centered	0.619	0.051	.746	12.113	<.001		
2	(Constant)	93.957	1.316		71.392	<.001		
	PPVT pre centered	0.400	0.096	.482	4.159	<.001	2.93	<.01
	Treatment	4.196	1.619	.155	2.592	.011	1.83	.07
	Interaction	0.281	0.112	.289	2.503	.014	1.76	.08

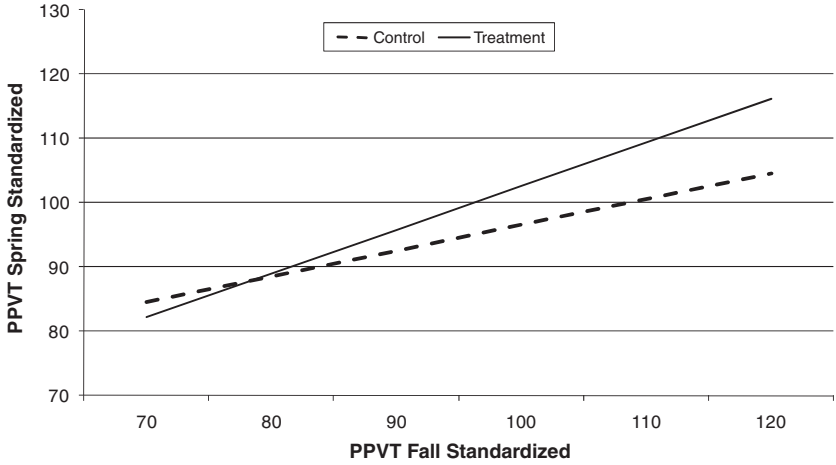

Figure 2. Regression analysis for the Peabody Picture Vocabulary Test (PPVT).

were especially pronounced for students with higher initial PPVT scores. For the target word measure, the overall estimated treatment effect was 1.71. For students with a fall PPVT score of 85, the effect size was 1.06 standard deviation units whereas the estimated effect for students with a fall PPVT score of 115 was 2.44 standard deviation units. (See Table 8.)

The treatment group scored higher than the control group on the posttest administration of the PPVT (corrected $p = .07$) and there was an interaction between the fall PPVT score and treatment (corrected $p = .08$). These results are presented in Table 6 and displayed in Figure 2. Although the estimated treatment effect for spring PPVT was not statistically significant at the .05 level, the overall effect size of 0.60 can be considered substantively important. For students with a fall PPVT score of 85, the effect size was .14 standard deviation units whereas the effect size for students with a fall PPVT score of 115

Table 7. Regression analysis: Listening comprehension coefficients

		Unstandardized		Standardized			Corrected	
Model		B	SE	β	t	Sig.	t	Sig.
1	(Constant)	3.068	0.172		17.788	<.001		
	PPVT pre centered	0.084	0.011	.580	7.667	<.001		
2	(Constant)	2.653	0.293		9.040	<.001		
	PPVT pre centered	0.045	0.022	.313	2.090	.039	2.09	.04
	Treatment	0.584	0.360	.121	1.621	.108	1.62	.11
	Interaction	0.050	0.025	.297	1.988	.049	1.98	.05

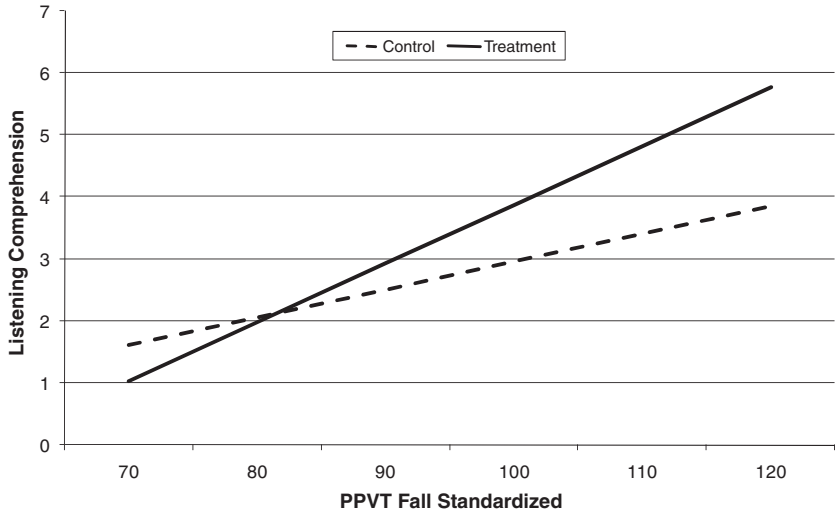

Figure 3. Regression analysis for listening comprehension. *Note*. PPVT = Peabody Picture Vocabulary Test.

was .81 standard deviation units. These effect size estimates indicate moderate treatment effects for students with average fall PPVT scores, large treatment effects for students with above average fall PPVT scores, and essentially no treatment effects for students with below average fall PPVT scores.

Table 7 and Figure 3 report the results of the regression analyses for listening comprehension. Overall, effects trended in favor of the treatment group (corrected $p = .11$) and there was an interaction between pretest PPVT scores and the treatment (corrected $p = .05$). Although the estimated treatment effect for spring PPVT was not statistically significant at the .05 level, the overall effect size of 0.42 can be considered substantively important. Again,

Table 8. Regression analysis: Metalinguistic awareness coefficients

		Unstandardized		Standardized			Corrected	
Model		B	SE	β	t	Sig.	t	Sig.
1	(Constant)	6.405	0.301		21.277	<.001		
	PPVT pre centered	0.122	0.019	.507	6.424	<.001		
2	(Constant)	5.640	0.516		10.936	<.001		
	PPVT pre centered	0.061	0.038	.256	1.627	.106	1.27	.20
	Treatment	1.075	0.632	.133	1.700	.092	1.33	.19
	Interaction	0.077	0.044	.276	1.759	.081	1.37	.17

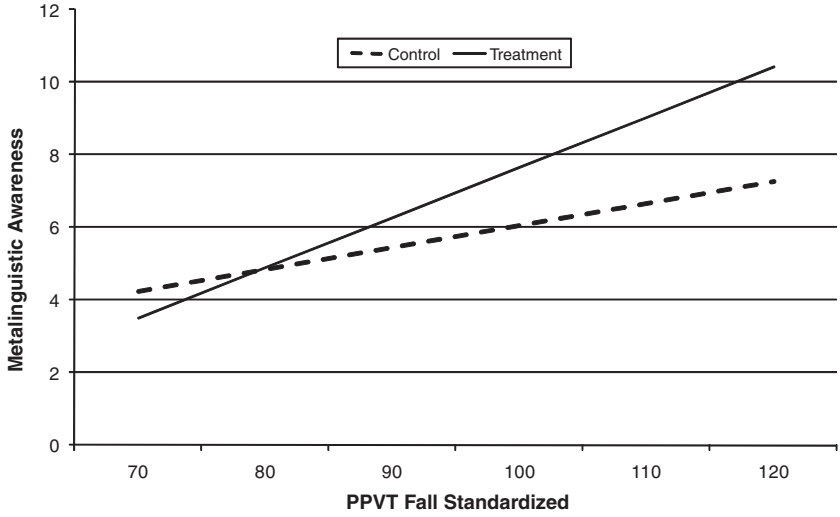

Figure 4. Regression analysis for metalinguistic awareness. *Note*. PPVT = Peabody Picture Vocabulary Test.

the treatment was more effective in increasing the listening comprehension scores of students with higher initial PPVT scores. The estimated treatment effect for students with a fall PPVT score of 85 was .08 standard deviation units, whereas the estimated treatment effect for students with a fall PPVT score of 115 was .74 standard deviation units.

Table 8 and Figure 4 report the results of the regression analyses for the measure of metalinguistic awareness. Overall, there were no statistically significant differences on this measure; however, effects trended in favor of the treatment group (corrected $p = .19$), and there was a trend toward an interaction between pretest PPVT scores and the treatment (corrected $p = .17$). For the metalinguistic measure, the estimated treatment effect for students with a fall

Table 9. Effect size measures at three different pretest PPVT standard scores (85, 100, and 115)

	Cohen's *d* (*SS* = 85)	Cohen's *d* (*SS* = 100)	Cohen's *d* (*SS* = 115)	Cohen's *d* (Overall)
Target word measure	1.06	1.75	2.44	1.71
PPVT posttest	0.14	0.48	0.81	0.60
Listening comprehension	0.08	0.41	0.74	0.42
Metalinguistic awareness	0.12	0.42	0.73	0.42

Note. PPVT = Peabody Picture Vocabulary Test.

PPVT score of 85 was .12 standard deviation units, whereas the estimated treatment effect for students with a fall PPVT score of 115 was .73 standard deviation units. Again, these effect sizes show essentially no benefit for students with below-average PPVT scores but moderate benefit for students with above average PPVT scores.

DISCUSSION

The purpose of this study was to evaluate the impact of an 18-week program of direct vocabulary instruction in kindergarten on measures of target word knowledge as well as transfer measures of generalized language and literacy. A second purpose was to examine whether students' initial receptive vocabulary knowledge moderated the effects of the vocabulary intervention.

Target Word Learning

Our findings indicated that kindergarten students who received vocabulary instruction demonstrated greater knowledge of target words at posttest compared with students who did not receive vocabulary instruction. Differences in target word learning between treatment and control students were characterized by large effect sizes. This finding is consistent with a growing body of research that suggests that young students can learn the meanings of sophisticated vocabulary through instruction that is direct and extended (Beck & McKeown, 2007; Biemiller & Boote, 2006; Coyne et al., 2009; Silverman, 2007).

Transfer Effects

Although we were interested in target word learning, it is not surprising that students learned words that were taught to them. Therefore, a central purpose of the study was to investigate whether vocabulary instruction would lead to transfer effects beyond learning target words. Studies have examined the transfer effects of vocabulary instruction with students in upper elementary school and beyond. Very few studies conducted with younger students, however, have included generalized measures of language and literacy. Although some of our findings bordered on statistical significance, we found a clear trend toward evidence of transfer to generalized measures.

We found moderate mean effect sizes favoring the treatment group on an experimenter developed measure of listening comprehension. This finding suggests that students who received vocabulary instruction were better able to answer questions about a story that contained taught vocabulary words. We also found moderate effect sizes on a standardized measure of overall

receptive vocabulary knowledge (PPVT–III). Although results were above our Type I error rate ($p < .10$), we also found moderate effects on an experimenter developed measure of metalinguistic awareness that required students to infer the meanings of unknown words.

The impact of our vocabulary intervention was similar to findings from a recent meta-analysis conducted by Elleman and her colleagues (2009). They found an average effect size of .50 for the impact of vocabulary instruction on custom measures of reading comprehension and an effect of .29 on measures of general vocabulary knowledge that is consistent with our results. However, only three of the studies in the meta-analysis included participants in Grades 2 and below, and the findings of these studies were mixed. Although the results of our study cannot be considered conclusive, they suggest that direct vocabulary instruction with younger students may produce similar effects on experimental developed measures of comprehension and generalized vocabulary knowledge as instruction with older students.

Our findings suggest that direct vocabulary instruction that is effective in teaching students the meaning of new words may also support the development of more generalized aspects of language and literacy. Although it is difficult to know the exact mechanism to explain these transfer effects, our results suggest a number of possibilities. First, our findings provide support for the instrumentalist hypothesis (Anderson & Freebody, 1981; Stahl, 1991), which holds that vocabulary learning directly leads to gains in comprehension, especially in text that includes a high percentage of taught vocabulary words. Whereas most studies that have examined the impact of vocabulary instruction on reading comprehension of older students, our findings indicate that this effect may also hold true for listening comprehension with younger students.

Second, our findings suggest that vocabulary instruction may enable students to become more successful independent word learners and help them strengthen their existing word knowledge. By receiving multiple and extended opportunities to manipulate language and interact with words deeply within engaging stories and other supportive contexts, students may have developed higher levels of metalinguistic awareness. The trend toward a heightened ability of students in the treatment group to infer the meanings of unknown words at posttest is consistent with this explanation. Increased metalinguistic awareness may have allowed students take advantage of other more informal and independent word learning opportunities during the course of the school year (Nagy, 2007) that resulted in higher levels of overall vocabulary knowledge as measured by the PPVT at posttest. The increased ability to reflect on and manipulate language, in this case words and their meanings, may have also allowed students to better understand assessment tasks and to demonstrate their existing knowledge in more sophisticated ways (Van Kleeck et al., 2003). Finally, by incorporating new words into their existing lexicons, students may have been able to consolidate and refine larger networks of semantic connections, thereby strengthening the representations of other partially known words which is

consistent with connectionist models of vocabulary learning (Landauer & Dumais, 1997).

Differential Responsiveness

We were also interested in whether the vocabulary intervention would be differentially effective—whether some students would respond more strongly to instruction than others. In our previous research (Coyne et al., 2009; Coyne et al., 2007), we found that overall receptive vocabulary measured prior to the start of the intervention was highly predictive of target word learning during the course of the instruction (see also Penno et al., 2002). The results of this study both replicate and extend these earlier findings.

We found that initial receptive vocabulary measured by the PPVT was strongly related to posttest performance not only on target word measures but also on all our other transfer measures. Students who had higher levels of overall receptive vocabulary knowledge at pretest benefitted more from the intervention than students with lower levels of vocabulary knowledge in comparison to students who did not receive the intervention. This finding was similar across all measures but was particularly pronounced on the transfer measures. For example, the impact of the intervention on transfer measures was characterized by large effect sizes (i.e., .73–.81) for students who had above-average vocabulary knowledge (i.e., standard score of 115) but by negligible effect sizes (i.e., .08–.14) for students who had below-average vocabulary knowledge (i.e., standard score of 85).

These findings provide an illustration of the Matthew Effect in vocabulary learning (i.e., the rich get richer while the poor get poorer; Stanovich, 1986). It is likely that students with more highly developed lexicons were better able to take advantage of existing vocabulary knowledge and perhaps higher levels of general background knowledge to support new word learning.

IMPLICATIONS

Increasingly, schools are recognizing the importance of supporting students' vocabulary development in the early grades. Indeed, schools are beginning to consider early vocabulary instruction with a growing sense of urgency because of the strong relationship between early language development and later reading outcomes and the increasing number of student with diverse learning needs entering kindergarten with limited exposure and experiences with language and literacy.

A growing body of evidence suggests that direct and extended vocabulary instruction enables young students to learn the meanings of sophisticated vocabulary (Beck & McKeown, 2007; Biemiller & Boote, 2006). Although there

are strong arguments that learning the meanings of individual words is important in itself (e.g., Biemiller, 2001), many researchers and practitioners hold an implicit hypothesis that vocabulary instruction affords additional benefits beyond learning target words. The results of this study suggest that this may indeed be the case and provide schools and teachers with a stronger rationale for incorporating vocabulary instruction into their daily practice.

We found that students with higher levels of vocabulary knowledge prior to instruction, however, responded more strongly to instruction than their peers with less developed vocabularies. This finding has a number of implications for both researchers and practitioners. First, work should continue to make classroom vocabulary instruction more supportive and efficacious for students with lower levels of vocabulary knowledge. Although we still don't know the ideal components to maximize the effectiveness of vocabulary instruction for these students, possible areas for improvement could include making more informed, and perhaps individualized, choices about words targeted for instruction, increasing learning opportunities, incorporating additional review and reinforcement, and carefully aligning instruction to students' existing word and world knowledge.

Even with carefully designed and highly differentiated classroom vocabulary instruction, however, teachers and practitioners should not necessarily expect instruction to be equally effective for all students. This suggests that schools should also consider identifying students who may not respond to classroom vocabulary instruction and increase the intensity of support for those students. Because of the need to intensify instruction for some students, it may be helpful to think about supporting vocabulary development within a multitier approach to instruction and intervention (Loftus et al., in press). A multitier or response-to-intervention approach has been used with promising results in other areas of literacy such as beginning reading instruction (Gersten et al., 2008). There are number of challenges, however, to conceptualizing supporting vocabulary development within a response-to-intervention framework. These include developing efficient measures for screening and monitoring progress (Catts et al., 2005) as well as conceptualizing and developing supplemental interventions that accelerate student learning. To date, there is little research that can help schools implement a multitier system of vocabulary instruction and intervention.

LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

A limitation of this study is its quasi-experimental design. Because of size and power constraints, we combined analyses across both whole class and small group instruction and we were unable to fully model the nested structure of the data. In addition, after we corrected for clustering, some of our results only bordered on statistical significance. Therefore, our findings should be considered

exploratory and must be interpreted cautiously. Future studies should use more rigorous experimental designs that take into account instructional context and the nested nature of the data.

Pearson, Hiebert, and Kamil (2007) recently wrote that the field's current ability to assess the impact of vocabulary instruction is "grossly undernourished, both in its theoretical and practical aspects" (p. 282). We have also found assessing vocabulary learning to be challenging. For example, far transfer measures of generalized vocabulary knowledge and comprehension are often not sensitive enough to capture the effects of vocabulary instruction and there are few options for assessing near transfer (Elleman et al., 2009; Pearson et al., 2007). Therefore, we used experimenter-designed measures of word learning and metalinguistic awareness and an experimenter modified measure of localized listening comprehension. The psychometric properties of these measures were clearly less strong than standardized measure of language and literacy. It will be important for researchers to continue to work to identify and develop technically adequate measures to assess both the proximal and distal impacts of vocabulary instruction.

A longstanding goal for both researchers and practitioners has been to work to close the vocabulary gap among students. A troubling finding of this study was that our vocabulary instruction benefitted some students more than others. This finding suggests that instruction widened the gap rather than decreased it. A critical direction for vocabulary researchers, and one that we are currently pursuing, is designing and evaluating interventions that significantly increase the intensity of vocabulary support for students at risk of experiencing language and learning difficulties within a comprehensive and coordinated system of multitier supports. Closing the vocabulary gap will continue to be an extremely challenging but incredibly urgent problem for schools and researchers alike.

REFERENCES

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Anderson, R. C., & Nagy, W. E. (1992). The vocabulary conundrum. *American Educator*, *16*(4), 14–18, 44–47.

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

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

Becker, W. C. (1977). Teaching reading and language to the disadvantaged: What we have learned from field research. *Harvard Educational Review*, *47*, 518–543.

Biemiller, A. (2001). Teaching vocabulary: Early, direct, and sequential. *American Educator*, *25*, 24–29.

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

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, *93*, 498–520.

Brabham, E. G., & Lynch-Brown, C. (2002). Effects of teachers' reading aloud styles on vocabulary acquisition and comprehension of students in the early elementary grades. *Journal of Educational Psychology*, *94*, 465–473.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Catts, H. W., Hogan, T. P., & Adolf, S. M. (2005). Developmental changes in reading and reading disabilities. In H. W. Catts & A. G. Kahmi (Eds.), *The connections between language and reading disabilities* (pp. 25–40). Mahwah, NJ: Erlbaum.

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

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

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoomiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality*, *12*, 145–162.

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator*, *22*(1–2), 8–15.

Dale, E., & O'Rourke, J. P. (1981). *The living word vocabulary*. Elgin, IL: Dome.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test–III*. Circle Pines, MN: American Guidance Service.

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

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*, 174–187.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide* (NCEE 2009–4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved March 3, 2009, from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

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

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice*, *15*, 198–205.

Justice, L. M., 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.

Kimer-Simon, K. L. (2003). *Storybook activities for improving language: Effects on language and literacy outcomes in Head Start preschool classrooms*. Unpublished doctoral dissertation, University of Oregon, Eugene.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's Problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211–240.

Lockett, J. N., & Shore, W. J. (2003). A narwhal is an animal: Partial word knowledge biases adults' decisions. *Journal of Psycholinguistic Research*, *32*, 477–496.

Loftus, S., Coyne, M. D., McCoach, D. B., Zipoli, R., Kapp, S., & Pullen, P. (in press). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at-risk for language and literacy difficulties. *Learning Disabilities Research & Practice*.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*, 86–92.

Nagy, W. E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52–77). New York: Guilford.

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: Reports of the subgroups*. Bethesda, MD: National Institute of Child Health and Human Development.

Nicholson, T., & Whyte, B. (1992). Matthew effects in learning new words while listening to stories. In C. K. Kinzer & D. J. Leu (Eds.), *Literacy research, theory, and practice: Views from many perspectives: Forty-first yearbook of the National Reading Conference* (pp. 499–503). Chicago: National Reading Conference.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, *42*, 282–296.

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew Effect? *Journal of Educational Psychology*, *94*, 23–33.

Puhalla, E. M. (2005). *Teaching vocabulary from narrative and information text: Examining the effects of instructional intensity and judicious review on the vocabulary and expressive language performance of first-grade children at-risk of early reading difficulties*. Unpublished doctoral dissertation, Lehigh University, Bethlehem, Pennsylvania.

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

Salvia, J., & Ysseldyke, J. E. (Eds.). (1998). *Assessment* (7th ed.). Boston: Houghton Mifflin.

Scarborough, H. (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.

Senechal, M., Thomas, E., & Monker, J. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology*, *87*, 218–229.

Share, D. L., & Leiken, M. (2004). Language impairment at school entry and later reading disability: Connections at lexical versus supralexical levels of reading. *Scientific Studies of Reading*, *8*, 87–110.

Shefelbine, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior*, *22*(1), 71–97.

Shore, W. J., & Kempe (1999). The role of sentence context in accessing partial knowledge of word meanings. *Journal of Psycholinguistic Research*, *28*, 145–163.

Silverman, R. D. (2007). Vocabulary development of English-language and English-only learners in kindergarten. *The Elementary School Journal*, *107*, 365–383.

Stahl, S. A. (1991). Beyond the instrumentalist hypothesis: Some relationships between word meanings and comprehension. In P. J. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 157–186). Hillsdale, NJ: Erlbaum.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, *56*, 72–110.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–406.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947.

Strong, C. (1998). *Strong Narrative Assessment Procedure*. Eau Claire, WI: Thinking Publications.

Van Kleeck, A., Stahl, S. S., & Bauer, E. B. (2003). *On reading books to children: Parents and teachers*. Mahwah, NJ: Erlbaum.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, *11*, 3–32.

What Works Clearinghouse. (2008). *Procedures and standards handbook, version 2.0*. [Electronic version]. Retrieved February 15, 2009, from http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf

Zipoli, R., Coyne, M. D., & McCoach, D. B. (in press). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically-related and embedded word review. *Remedial & Special Education*.

APPENDIX

Table A1. Storybooks and target vocabulary

Author	Title and Publisher	Target Vocabulary
Helen Lester (1990)	*Tacky the Penguin*. Boston: Houghton Mifflin.	peculiar, apex, plummet
Helen Lester (1989)	*A Porcupine Named Fluffy*. Boston: Houghton Mifflin.	slender, drenched, sprawl
James Marshall (1998)	*Goldilocks and the Three Bears*. New York: Puffin Books.	cycling, scalding, parlor
James Marshall (1996)	*The Three Little Pigs.* New York: Puffin Books.	sturdy, weep, festival
Elizabeth Lane (2003)	*Anansi and the Seven Yam Hill*. Menlo Park, CA: Electronic Education.	irritated, plot, exclaim
Wendi J. Silvano (2003)	*Lorenzo's Llama*. Sunnyvale, CA: Electronic Education.	slope, regretful, sprinted
Ezra Jack Keats (1976)	*A Snowy Day*. New York: Puffin Books.	grin, saunter, glum
Rosemary Wells (2005)	*McDuff Moves In*. New York: Hyperion.	terrace, thrash, serene
Rosemary Wells (2001)	*Yoko's Paper Cranes*. New York: Hyperion.	voyage, gaze, soar
Maryn Roos (2000)	*The City Mouse and the Country Mouse*. Sunnyvale, CA: Electronic Education.	cautious, devour, metropolis
Faith Ringgold (1996)	*Tar Beach*. New York: Dragonfly Books.	boulevard, hoist, immense
Ho Baek Lee (2003)	*While We Were Out*. La Jolla, CA: Kane/Miller Book Publishers.	grasp, minute, weary
David Shannon (2002)	*Duck on a Bike*. New York: Blue Sky Press.	courageous, halt, terrified
Rosemary Wells (2000)	*Bunny Cakes.* New York: Puffin Books.	confused, crouched, furious
Jill Murphy (1999)	*Five Minutes Peace*. New York: Penguin Putnam Books.	residence, stalk, stout
Sally Hunter (1999)	*Humphrey's Corner*. New York: Henry Holt and Company.	ascend, elated, vivid
W. Nikola-Lisa (1991)	*Night is Coming*. New York: Scholastic Inc.	brook, cavern, meadow
Laura Appleton-Smith (1997)	*The Sunset Pond*. Lyme, NH: Flyleaf Publishing.	considerate, dusk, vanished