Effects of vocabulary instruction on measures of language processing: Comparing two approaches

Article in Early Childhood Research Quarterly - December 2014
DOI: 10.1016/j.ecresq.2014.06.002

2 authors:

Margaret G. Mckeeown
University of Pittsburgh
53 PUBLICATIONS 3,243 CITATIONS

Isabel Beck
University of Pittsburgh
74 PUBLICATIONS 3,976 CITATIONS
Effects of vocabulary instruction on measures of language processing: Comparing two approaches

Margaret G. McKeown *, Isabel L. Beck

University of Pittsburgh, United States

A R T I C L E   I N F O

Article history:
Received 13 August 2013
Received in revised form 15 May 2014
Accepted 2 June 2014
Available online 11 June 2014

Keywords:
Vocabulary
Comprehension
Read-alouds
Kindergarten

A B S T R A C T

This study examined effects on kindergartners (n = 131) of two approaches to vocabulary instruction, repetition and interactive, and a control condition, along a progression of language processing, using a within subject design. The repetition condition featured repeated readings of a story and practice with definitions. The interactive condition featured multiple contexts and active processing of the words. Students were assessed with experimenter-designed measures of meaning recognition, comprehension, and production. Repetition and interactive approaches enabled students’ recognition of word meanings and higher-order processing compared to the control. Two measures of higher-order processing showed advantage for interactive instruction relative to repetition instruction.

© 2014 Elsevier Inc. All rights reserved.

There no longer seems to be much controversy in the idea that attention to vocabulary needs to begin early in children’s schooling, as early as kindergarten, and perhaps even as early as preschool. Evidence of the gap in vocabulary for children of different SES groups by age three (Hart & Risley, 1995) and consistent findings that children’s early vocabulary knowledge strongly predicts their later reading success (Biemiller & Slonim, 2001; Cunningham & Stanovich, 1998; Hart & Risley, 1995; Wagner et al., 1997) demonstrate the urgency of getting young students off to an early start. Consensus reports on literacy have universally recommended early attention to children’s vocabulary development in school (Common Core State Standards, 2010; National Early Literacy Panel, 2008; National Reading Panel, 2000; Snow, Burns & Griffith, 1998).

Questions remain, however, about whether vocabulary instruction for young children can be successful in promoting not only word knowledge but also children’s literacy potential. Because the value of vocabulary instruction lies in its ability to affect students’ literacy development, an important goal of this study was to gauge the impact of instruction not only on word knowledge but also on comprehension.

Empirical and theoretical foundation

A traditional vehicle for enhancing children’s language skills, including vocabulary, has been storybook reading. Positive correlations between being read to and eventual reading achievement have appeared in journals for over 50 years (Teale, 2003). Positive effects specifically on vocabulary were documented by Mol, Bus, and De Jong (2009) in a meta-analysis of 31 studies of classroom interactive read alouds. Using read aloud events to promote vocabulary learning aligns with a language development perspective, which holds that exposure to and interaction with language in natural settings is the key to literacy development.

With the literature pointing to reading aloud as a natural vehicle for enhancing vocabulary, researchers began to focus more directly on the effects of single read-aloud events on incidental learning of specific vocabulary items that appeared in the text (Biemiller & Boote, 2006; Elley, 1989; Nicholson & Whyte, 1992; Penno, Wilkinson, & Moore, 2002; Robbins & Ehri, 1994; Sénéchal, Thomas, & Monker, 1995). Researchers generally found learning effects, but they were quite limited, for example, no gain in Justice, Meir, and Walpole (2005) to 3% and 15% on two different stories in Elley (1989). Studies employing read alouds for vocabulary development then moved from simple story readings to enhancements such as repeated readings of stories, assuring that the contexts supported word meanings (Robbins & Ehri, 1994), and explaining word meanings as a story was read (Biemiller & Boote, 2006; Collins, 2010; Elley, 1989; Justice et al., 2005; Penno et al., 2002).

Repeated readings of stories and providing word meaning brought substantial increases in word learning. These instructional
enhancements chiefly utilized memory and association processes, helping students associate a word with its meaning and boosting the chances that the association would be remembered by offering repeated encounters. The theoretical perspective underlying this approach is based on the hypothesis that an exposure to a word in context establishes an initial referent for the word, and direct explanation of the word’s meaning provides a generalization of word meaning that includes sufficient information for comprehending a variety of uses and contexts for the word. Repetitions of readings and explanations strengthen the connection to allow future incidental encounters with the new word to be better understood (Biemiller & Boote, 2006).

Another approach to enhancing students’ vocabulary development draws from a cognitive processing framework, which entails the theoretical notion that active processing—active or attentive mental manipulation of ideas—is necessary to learning if the learner is to achieve the ability to use and apply new information (Brown, Bransford, Ferrara, & Campione, 1983; Miller, 2003; Sternberg, 1979, 1982). Implications of a cognitive-processing perspective for vocabulary center on the need for learners to interact with and integrate various specific contexts of word use in order to form generalizations that are of sufficient quality to assist comprehension (Bolger, Balass, Landen, & Perfetti, 2008; Nagy & Scott, 2000; Perfetti & Stafura, 2014; Perfetti & Hart, 2002).

Two early reviews of vocabulary instructional research supported a cognitive processing orientation. Both Mezynski (1983) and Stahl and Fairbanks (1986) concluded that instruction that affected comprehension included multiple exposures to each word, both definitional and contextual information, and active, or deep, processing. More recent work has elaborated the view that semantic learning requires multiple exposures in a variety of contexts and active processing (Marulis & Neuman, 2013; Nagy & Scott, 2000; Perfetti & Stafura, 2014; Perfetti & Hart, 2002). As Perfetti and Hart’s (2002; Perfetti, 2007) lexical quality hypothesis posits, experiencing words in multiple, informative contexts allows learners to build rich networks of connections that in turn lead to complex, flexible, and nuanced representations of word meaning. Such representations enable learners to bring the most relevant ideas to bear in making sense of subsequent contexts in which the word is met.

A cognitive processing orientation underlies several instructional techniques for vocabulary. For example, semantic features analysis (Anders, Bos, & Filip, 1984) and semantic mapping (Margosein, Pascarella, & Pfaum, 1982) engage learners’ processing by having students examine how words are related through analysis and discussion of word characteristics (Johnson & Pearson, 1978, 1984). Both techniques have resulted in improved word learning and comprehension (Anders et al., 1984; Margosein et al., 1982).

Studies by Beck and McKeown and their colleagues with fourth grade students also reflected a cognitive processing orientation (Beck, Perfetti, & McKeown, 1982; McKeown, Beck, Omanson, & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985). Instruction was designed to promote fluent access to word meanings and rich semantic connections through multiple exposures of target words in different contexts and activities to engage student processing. For example, students were asked to discuss if they would berate someone who had inspired them. Such a question prompts students to consider the meaning of the two words, activate the circumstances under which berating is relevant, and decide whether those circumstances fit a person who inspires one. The instruction was found to affect not only knowledge of word meanings, but also students’ higher-order language processing, including speed of access to taught words, ability to integrate new words into contexts, and comprehension of text that used the taught words.

Vocabulary research featuring active processing initially involved students in intermediate grades through high school. But many researchers have now used a cognitive processing framework to develop vocabulary instruction for much younger students (Beck & McKeown, 2007; Coyne, McCoach, Lofius, Zipoli, & Kapp, 2009; Coyne, Simmons, Kame’enui, & Stooanimal, 2004; Silverman, 2007; Wasik & Bond, 2001). Based on theory and prior research with older students, we might expect such studies with younger students to impact higher-order abilities such as comprehension as well. However, although these studies have found positive impact on word learning, they have not examined effects on comprehension, with one recent exception: the study by Coyne et al. (2010), which showed a trend (p = .11) toward enhanced comprehension on a listening measure for kindergartners who learned vocabulary through active processing instruction relative to no treatment controls.

**Purpose**

This study has three major purposes. The first is to compare two types of vocabulary instruction for kindergarten children. The second is to examine the effects of vocabulary instruction along a progression of language processing, from recognition of word meaning to comprehension and production, and the third is to use innovative measures to examine points along this processing progression.

**Rationale for the study**

There has been no research comparing the two types of instruction considered here, which are designed around either repeated story readings or cognitive processing activities. Both begin with read alouds and provide definitions for target words. Repeated readings instruction comprises repeated readings of stories and practice of target word definitions. Cognitive processing instruction features presentation of a variety of contexts for the words and asking students to generate, evaluate, and explain various uses for the words. Such an investigation is valuable, as both approaches are prevalent in the instructional and research literature. Because the two approaches engage different processes, understanding the effects of each sheds light on the kind of processing that may promote different aspects of verbal functioning.

**Rationale for outcomes measured**

Assessment of what students learn from vocabulary instruction has been quite restricted (Pearson, Hiebert, & Kamil, 2007). Typically, studies employ measures of word knowledge that involve, essentially, knowing a definition for the words. Results of such assessments are limited in that we do not know if they indicate that students have truly developed a generalized representation of what a word means, or if students are simply recalling a definition from instruction. Some studies also attempt to measure comprehension effects, often using standardized measures of text comprehension, usually with meager results (Ellemor, Lindo, Morphy, & Compton, 2009). A definition task and a general comprehension measure represent extreme ends of a continuum from proximal to distal outcomes. Thus, for example, positive results on a definition task and null results on a comprehension task would leave a wide gap in understanding what students can do with the words they have been taught.

A key goal of this study was to gauge the impact of instruction on a progression of language processes, from recognition of word meanings, which calls on lower-order processes of memory and association, to higher-order processes of comprehension and production. Higher-order processes go beyond lower-order processes such as perception, memory, and association, by requiring combinations of cognitive actions and use of information beyond what is given.
Two aspects of comprehension were measured, context integration and text comprehension. Context integration is the ability to access relevant meanings of words encountered in context and fit those meanings together in a way that allows one to make sense of the context so that comprehension can proceed. For example, in the context “After the contest winners were announced, Sara ran up to console Meg,” the reader needs to use knowledge of console to generate the implication of the context, that if Sara is consoling Meg, that means Meg must have lost the context.

Context integration was initially noted as an important component of comprehension by Jenkins, Pany, and Schreck (1978) and Kame’enui, Carnine, and Freschi (1982). More recently, studies by Perfetti and his colleagues (Perfetti, 2007) have shown that skilled comprehension is marked by more successful integration of meaning. In a study with fourth-graders that compared effects of instruction based on definitions to instruction based on active processing, McKeown et al. (1985) presented contexts such as the Meg scenario presented above followed by questions such as: “How do you think Meg did in the contest?” The active processing instruction group was significantly more able to respond to the questions.

Comprehension of connected text is a complex processing event that calls on vocabulary knowledge, among other abilities. Successful comprehension of a text is generally indicative of high-level knowledge of the vocabulary within the passage. Thus texts containing instructed words have been used to measure comprehension effects of instruction (Beck et al., 1982; Coyne et al., 2010; Kame’enui et al., 1982; McKeown et al., 1985; Pany & Jenkins, 1978).

Another aspect of language processing examined in this study is the ability to use words spontaneously to express meaning. Expressive, or productive, vocabulary represents higher-order processing relative to recognizing word meanings (Stahl, 1986), as it indicates word knowledge that is of sufficient quality to make the words accessible for production.

Method

The study involved developing, implementing, and comparing instruction for kindergartners for two approaches to vocabulary development, and a control condition: repetition, interactive, and control. The design was within subject, with each student experiencing each kind of instruction. The design treated each student as his/her own control, thus reducing error variance and increasing statistical power.

The repetition condition was based on Biemiller and Boote’s repeated readings approach (2006), which comprises repeated readings of a storybook with explanation of target word meanings during reading and activities that review word meaning. The interactive condition was based on a cognitive processing approach, as represented in work by Beck et al. (1982) and Coyne et al. (2010). It provides one reading of a story and then focuses on interactions around word meanings in which children are invited to think about and respond to words. Both approaches provide a story context, provide explicit definitions for words selected from the story, and follow-up activities. However, the story context and the definitions are the only content of the activities in the repeated readings method. In contrast, the cognitive processing-based instruction offers additional contexts for the words and engages students in responding to the contexts and generating their own contexts.

The control condition comprised reading a story aloud one time, which is a customary way of using read alouds in classrooms (Beck & McKeown, 2007; Hoffman, Roser, & Battle, 1993; Lickteig & Russell, 1993). The control condition represented the opportunity to encounter the words incidentally in a meaningful context. We chose this as a comparison condition because it enabled us to know that students had indeed heard the words, providing some opportunity to connect them with meaning. This allowed us to explore how such incidental encounters compared to learning from direct instruction.

The following hypotheses underlay the study:

1. Repetition and interactive conditions will be more effective than the control condition for promoting children’s recognition of word meanings, integration of words and context, listening comprehension, and spontaneous production of target words.
2. Repetition and interactive conditions will not differ in the extent to which they promote children’s recognition of word meanings.
3. The interactive condition will show advantage over the repetition and control conditions for higher-order language processing (context integration, listening comprehension, and production).

Participants

Participants were 131 kindergarten students (67 boys and 64 girls) and their teachers in eight classrooms from a public school district in the northeast. The number of students per classroom ranged from 14 to 18. The school district population represents four working-class communities, and approximately 50% of the students receive free or reduced-price lunch. Approximately 75% of the students were European-American and 25% were African-American. All students spoke English as their first language. The Peabody Picture Vocabulary Test (PPVT-4) (Dunn & Dunn, 2007), which was given to all students for descriptive purposes, showed a mean of 101 and a standard deviation of 19. There was a wide range, with a minimum of nine and a maximum of 154.

All eight kindergarten teachers in the district, housed at an Early Learning Center, were invited to participate in the study and all agreed to do so. All teachers were white females. The number of years of teaching experience ranged from two years to 25 years. All teachers had spent their entire teaching careers in the district in which the study occurred. The teachers each received a $300 stipend for their participation.

Materials

Materials for the study were based on three high-quality children’s tradebooks. Our identification of high-quality texts focused on texts whose language and events were intellectually challenging, providing sufficient gist to support children in using and responding to language. This included story plots that exhibited some complexities of events and subtleties in expressing ideas, and inclusion of words beyond familiar, everyday vocabulary. We also sought texts in which the linguistic content was primary, that is, communicating the story did not rely heavily on the pictures (Beck & McKeown, 2001). We identified 10 words to teach from each story, for a total of 30 words. The three stories were: Mr. Tumor’s Ties (Coca-Lefler, 1999), Mrs. Potter’s Pig (Root, 1996) and A Pocket for Corduroy (Freeman, 1978).

Target words. The instructed words can be described as Tier-2 words, meaning that they fall into a level of utility in the language between everyday conversational words and low frequency words that apply to specific content domains (Beck, McKeown, & Omanson, 1987). Tier-2 words have a great deal of overlap with academic words (Coxhead, 2000) in that they are found across content domains and are more frequently found in text than in oral language. They are general utility words that label concepts that even young children are likely familiar with. For example, reluctant and ponder are Tier-2 words that young learners are unlikely to know. But young learners do understand the concepts that the words represent; for reluctant, of not particularly wanting to do something, such as go to the dentist; and for ponder, thinking things over, such as which flavor ice cream to choose.
Table 1
Target words from each story.

<table>
<thead>
<tr>
<th>Story</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mrs. Potter's Pig</td>
<td>glee, shriveled, clutch, curious, stunned, perfectionist, strolled, devoured, inseparable, plead, reluctant, insisted, affectionate, patiently, precious, distraught, rummage, ponder, inspired, ponder, eager</td>
</tr>
<tr>
<td>Pocket for Corduroy</td>
<td>appropriate, charming, stern, concentrate, admire, gazed, perplexed, spectacular, timidly, abruptly</td>
</tr>
<tr>
<td>Mr. Tanen's Ties</td>
<td></td>
</tr>
</tbody>
</table>

We examined each story for the presence of Tier-2 words and, to insure that we were able to teach 10 words for each story, identified places in the stories where Tier-2 words could be incorporated. We added words by substituting synonyms, for example, abruptly for suddenly, or by adding a word to describe some implied idea. For example, Mr. Tanen in Mr. Tanen's Ties had an extensive and interesting collection of ties, and we added the word spectacular to describe it. Table 1 presents the target words from each story.

For each target word, we designed a friendly explanation of the word's meaning, and activities with which students engaged. Explanations of word meaning were based on Beck, McKeown, and Kucan's (2002) student-friendly explanations model, in which word meanings are presented in everyday language. For example, admire was explained as: “If you admire someone, you think they are really special because of something they have or do.”

Design of instructional activities. The repetition and interactive conditions provided 12 encounters for each word, with an encounter defined as an opportunity to connect a word to its definition or to a context, including the word's appearance in the story context or in one of the instructional activities. Twelve was selected because that number of encounters had affected comprehension in a study with fourth-graders (McKeown et al., 1985). To accomplish this number of encounters, a seven-day instructional cycle was planned for each of the three stories that included a pattern of encounters for each word. The activities for each day of a cycle were also held constant across cycles within the two experimental conditions. A version of each of the three story cycles was developed for each instructional condition.

Lessons in all three conditions began with a story read aloud; hearing the words read in the context of the story was counted as one encounter. That was the only encounter for the control condition. All lessons were taught by the classroom teacher.

Repetition condition. In the repetition condition, each time a target word was encountered in the story, the teacher stopped and read the friendly explanation for the word. After story reading on the first day of instruction, five of the ten targeted words were presented by referring to the context from the story in which each had occurred, paraphrasing the context, and presenting the friendly explanation. On the second day, the story was read and each definition presented as the word was encountered, and after the story, the other five words were presented as on Day 1. On the third day, again the story was read and each definition presented as the word was encountered. On Days 4 through 7, students were engaged in activities to practice the friendly definitions.

Activities in the repetition condition focused on learning and practicing the meanings of the targeted words. The specific activities included game-like formats such as “Concentration,” in which students turned over cards to match words with definitions, “Yes/No” in which students had to signal whether a word was paired with its correct definition, and “Stop” in which the teacher presented a word or definition and read down a list of definitions or words and students were to call “stop” when the correct match was reached. Lessons in the repetition condition concluded with an activity that asked students to choose which of two words matched a provided definition.

Interactive condition. In the interactive condition, the stories were read one time, without interruption. On the first instructional day, after the story was read, five of the words were introduced, with the initial introduction the same as in the repetition condition, by referring to the context from the story in which the word had occurred, paraphrasing the context, and presenting the friendly explanation.

Follow-up activities were designed to prompt students to think about and respond to situations with the taught words. Such activities were initiated after each word introduction with an activity in which students were asked to distinguish between examples and nonexamples of the word’s application. For instance, students were asked to say “reluctant” if the teacher said something they would be reluctant to do. The teacher then offered the following: “Holding a tarantula spider;” “Petting a kitten;” and “Jumping out of a tall tree.” On Day 2, the other five words were introduced in the same way and the first five words were reviewed.

Instruction in Days 3 through 7 included additional activities such as prompting students to make choices about the use of words and create and explain contexts for words. An example of making choices included judging whether sentences using target words made sense and explaining why/not, given sentences such as: “Chris and Michael never liked each other and were completely inseparable.” Creating and explaining contexts for words included responding to, for example, “Can you tell me a time when you need to be patient?” and “Why would you ponder your next move in checkers?”

Measures

To measure outcomes of vocabulary instruction along a progression of language processing, students were assessed with experimenter-designed measures of meaning recognition, context integration, listening comprehension of connected text, and production. Experimenter-designed measures were used because no standardized measures were appropriate to the outcomes targeted by the study. Our goal was to measure what students were able to do with the words they had been taught, from meaning recognition to comprehension and production. Standardized measures tap general skill, and it would have been highly unexpected to see differences on such measures after only seven days of instruction, which is what was delivered to each student for each condition. Table 2 summarizes the measures. In this section, we describe the role of each measure, its process of development, and how it was administered and scored.
Meaning recognition measure. The meaning recognition measure assessed an initial level of word knowledge that included recognizing a paraphrase of word meaning and typical use of the word.

Development. This measure followed a format developed by Beck and McKeown (2007) and used by other researchers (Aptorp et al., 2012; Coyne, McCoach, & Kapp, 2007; Coyne et al., 2009). The measure uses four questions for each of the 30 instructed words, two that ask about the definition and two that ask about use of the word in context, each with a true and a false item. For example, the items for insist are: Does insist mean that you had to do something? (definition—true). Does insist mean drawing a picture? (definition—false). If your mother made you clean up your room right now, would she be insisting? (context—true). If your teacher tells you she likes your jacket, is she insisting? (context—false). The pretest contained 120 items for the 30 target words and 10 items created for familiar words to make the test more accessible for young students.

The internal consistency reliability of the meaning recognition measure was calculated using the pretest scores, because all students had received all items at the same time prior to instruction. The analysis yielded a reliability coefficient (alpha) of .64. Measures using the same format, but with different words and number of items (80 per grade), with a much larger sample of students obtained coefficients ranging from .77 for kindergarten to .87 for grade four (Aptorp et al., 2012). Approximately 1100 students were tested per grade in that study.

Administration. The pretest was presented to students in whole-class format in five sessions across three days. Students marked an answer sheet by drawing an X over a smile face for yes answers and over a frown face for no answers. The procedure for the items was introduced, practiced, and discussed to check for children’s understanding for several practice items before the actual test began. The test was administered by one of the researchers and monitored by the classroom teacher to assure that students were responding to each item and that they were doing their own work. Posttests consisted of the items for the 10 words in the particular instructional cycle just completed, and thus contained 40 items each. Each posttest was presented to students in whole-class format in a single session.

Scoring. One point was awarded for each correct answer, for a total score of 40 for each cycle. The familiar items on the pretest were not scored.

Context integration measure. The context integration measure was developed to assess an aspect of comprehension, the ability to accommodate a word’s meaning within a sentence context in order to make sense of the context. The measure consisted of sentences with an instructed word, each followed by a question that asked students to infer implications of the context. An example is: “Sam was stunned when he looked into the doghouse. What do you think Sam saw?” For this item, students would have to access their understanding that stunned meant very surprised and then connect to something about a doghouse that might cause that reaction, such as, “his dog and four new puppies.”

The contexts in this task were developed to be situations familiar to children, and to be neutral or even shade toward the opposite emotion conveyed by the vocabulary word alone. The item about Sam being stunned seemed neutral, as someone might have any number of reactions to looking into a doghouse; the specific reaction is governed by accommodating stunned into one’s interpretation. An item that might suggest opposite emotions was: “Jim had to insist that Freddy go on the Merry-Go-Round. What did Freddy think about the merry-go-round?” Our hunch was that most young students would respond positively to a carnival ride, and thus being able to accommodate the concept of insisting was critical to correctly responding to the question.

Development. An item was developed for each word. We conducted several rounds of pilot testing. First, as we had never used this format with very young students, we tried items using only familiar words to see if five and six-year-olds could answer questions such as “Sam was surprised [rather than the target word stunned] when he looked into the doghouse. What do you think Sam saw?” We found that the children did well with the format. We then used target words and tested moderately high-ability children who were entering kindergarten and first grade, and found that they were generally able to respond to these items. The internal consistency reliability of the context integration measure was calculated using all students’ posttest scores. The analysis yielded a reliability coefficient (alpha) of .78.

Administration. Context integration measures were administered as posttests after each cycle of instruction. The measure was administered individually by a research staff member who recorded the student’s responses. To elicit maximum information from the students, the researcher followed up each question, asking how the student’s response fit with the word. So, for example, after a student responded to the item for stunned illustrated above, the researcher asked “How does that fit with being stunned?”

Scoring. Responses were awarded 1 point if they indicated understanding of the target word within the context. For example, for the item: “Mrs. Thomas was distraught when she looked at her garden. How did the garden look?” responses that were scored correct indicated that the garden did not look good—causing Mrs. Thomas to feel distraught. Sample correct responses included: “It was all messed up,” and “footprints and maybe animals sneaked into her yard.” Responses to the follow-up questions (“How does that fit with . . .?”) were used to help evaluate responses that seemed ambiguous or too sparse to enable a judgment of appropriateness.

Responses were scored by one team member and 20% were scored by a second team member to determine inter-rater reliability. For responses for all three cycles, agreement of 96% was reached (96%, 98%, 94% for each cycle respectively). Differences for the 20% were resolved through consensus.
Listening comprehension measure. The Listening Comprehension Measure consisted of stories for each cycle that contained target words, for which students were asked to recall and respond to questions.

Development. A story was developed for each cycle that used all 10 target words. The stories were similar in length, difficulty, and structure. All three were simple conflict-resolution narratives, contained two main characters, and were about two-thirds description and one-third dialogue. The number of words ranged from 232 to 239, and Flesch–Kincaid grade level readability ranged from 2.3 to 2.6. Ten questions were developed for each story. The questions were either framed around a target word or required a response that called upon the use of a target word in the story.

Administration. Each story was read aloud to students individually in separate sessions. After a story was read, the student was asked to recall the story and to answer 10 open-ended questions. The recall and question responses were audio recorded for later transcription.

Scoring. Recalls were scored for length and quality. Scoring for both dimensions involved dividing the stories and student recalls into content units, units of text about a clause in length (Omannson, 1982), which we have used in previous work (Beck et al., 1982; McKeown, Beck, & Blake, 2009). The level of importance of each unit was the basis for the quality measure. Importance was determined by categorizing each unit as major, support, or detail. Major units were assigned a value of three points, supporting units two points, and details one point, based on the expectation that recall indicative of successful comprehension will consist mostly of important units that make up the gist of the story. The story for Cycle 1 had 47 units, 6 main, 12 support, and 29 details for a possible total score of 71; the story for Cycle 2 had 43 units, 6 main, 11 support, and 26 details for a possible total score of 66; the story for Cycle 3 had 43 units, 6 main, 11 support, and 23 details for a possible total score of 63.

Recalls were scored by one team member and 20% were scored by a second team member to determine inter-rater reliability. Differences for the 20% were resolved through consensus. Across the three stories, 93% agreement was reached (91%, 95%, 94% for each story separately).

To score the questions, researchers met as a group to decide on acceptable answers. Question responses were then scored by one team member and 20% were scored by a second team member to determine inter-rater reliability. Differences for the 20% were resolved through consensus. Across the sets of questions for all three stories, 93% agreement was reached (94%, 100%, 85% for each story separately).

Production measure. The Production Measure for each cycle presented a picture of a scene designed to prompt use of the target words for that cycle.

Development. A picture was developed for each cycle’s set of words to depict an action or person that could be described by each word. Each picture included places and activities familiar to children and included both children and adults. For example, the picture for Cycle 1 was a sidewalk scene that included a boy who was grinning and waving his hands around, to represent being gleeful, two girls with their arms around each other to represent inseparable, and a woman clutching her purse.

Administration. Each production task was administered to students individually by a staff researcher in separate sessions. Each session was audio recorded for later transcription. At the beginning of the session, the researcher said, “I know you learned some new words this week. Let’s see, you learned: [named each target word]. Did you use any of these words when you were talking to people?” The researcher then explained that she would show a picture and ask the student to talk about it and then ask some questions about the people in the picture. For example, questions about the “stern” woman were: “What can you tell me about this lady? What can you tell by looking at her face?” Thus, students were reminded about the target words, but not directly asked about them in relation to the picture.

Scoring. A point was awarded for each target word that was used in talking about the picture.

Procedures

Consent forms were sent home with 139 students, and all 131 students who returned their forms participated in the study. The instructional conditions were treated as within-subject factors, allowing each student to be exposed to one cycle each of interactive instruction, repetition instruction, and the control condition.

Implementation. A workshop was conducted for the teachers before implementation began to explain the study in general terms, introduce our approach to vocabulary development, and present the lesson material. Teachers received a folder for each of the three stories that contained the storybook and scripted instructional activities for that story, tailored for their assigned condition. All teachers taught the same story during each cycle. The study consisted of three cycles, presented in the same order for all classes. The first cycle was based on Pocket for Corduroy (Freeman, 1978), the second on Mrs. Potter’s Pig (Root, 1996) and the third on Mr. Túmér’s Ties (Coca-Lefller, 1999). Each teacher taught each story and each condition. Thus, during each cycle, while students were instructed on the same story, they experienced different conditions depending on classroom.

Data collection. The meaning recognition measure was given as a pretest, and appropriate items from the pretest were used for the posttests for each cycle. All other measures were given as posttests. Because of the within-subject design, posttesting was conducted after the instruction for each cycle. Thus, words from each story were tested following the instruction associated with that story. The tests were administered in a fixed order: meaning recognition, context integration, listening comprehension, and production. Posttesting took place about five days after instruction ended depending on the day of the week instruction was completed.

Fidelity of implementation. During the intervention, we observed classrooms weekly and met with the teachers after observations to provide feedback as needed on the fidelity to the approach and to elicit teacher comments and concerns. From initial observations, it was judged that teachers were conducting the scripted lessons as designed. All lessons were audio recorded and subsequently transcribed, allowing us to conduct a systematic review for fidelity. We examined transcripts of lessons of teachers in the interactive and repetition conditions for one story, Pocket for Corduroy, which was the second story in the implementation. We checked whether each item in the lesson was presented, whether an appropriate student response had been obtained, and whether appropriate follow-up was then provided. Items corresponded to what the teacher was directed to say about a word meaning or use and the question students were asked about it. Each item, response, and follow-up was scored 0 if not present, 1 if present but not complete, and 2 if complete and appropriate. Fidelity for the repetition lessons was 100%, and for the interactive lessons averaged 99.7% (part of one question was omitted in one lesson).

Analytic plan

Data were analyzed to test our hypotheses that the interactive and repetition conditions would be more effective than control for meaning recognition, and that the interactive condition would be more effective than repetition and control for measures of comprehension and production. Because students within each classroom were exposed to each condition, the data for each condition can
be pooled across the eight classrooms. To account for within-subject nesting of vocabulary instruction and assessments, data were analyzed using linear mixed-effect model analyses of variance (ANOVA). Subject was treated as a random effect, and story cycle was treated as a repeated measure for all analyses of vocabulary assessments. Story and Condition had a consistent interaction effect so scores on all assessments were covaried by story. Post-hoc analyses of means were run using Least Significant Difference (LSD).

Results

Results are reported for the meaning recognition measure, the two measures of comprehension—content integration and listening comprehension of connected text—and the production measure. This set of measures was designed to tap a sequence of language processing.

Meaning recognition

For the meaning recognition measure, pre/post test as a repeated measure was added to the ANOVA. The analysis showed significant differences across conditions. Post-hoc analyses (LSD) revealed that the control condition differed significantly from both the repetition (p = .003; d = .35) and interactive (p < .001; d = .44) conditions, which did not differ significantly from each other. The results are presented in Table 3. Mean scores represent the growth in the number of correct items from pretest to posttest that students experienced within each condition of instruction.

Context integration

For the context integration measure, the ANOVA showed significant differences across conditions. Post-hoc analyses (LSD) revealed that the control condition differed significantly from both the repetition (p = .03; d = .27) and interactive (p < .001; d = .48) conditions. In addition, scores tended in favor of the interactive condition over the repetition (p = .07; d = .21). The results are presented in Table 4. Mean scores represent number of correct items out of the 10 items that students were tested on for each condition.

Listening comprehension

The separate ANOVA’s conducted for recall length and quality and for questions showed no differences among conditions. The results are shown in Tables 5 and 6. Mean scores for length represent number of units recalled for a story, and mean quality score represents the recalled units for a story weighted by level of importance.

Table 3
Means (SE) and analysis of variance results for recognition measure.

<table>
<thead>
<tr>
<th>Condition</th>
<th>M (SE)</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>4.92 (.45)</td>
<td>7.74</td>
<td>.001</td>
</tr>
<tr>
<td>Repetition</td>
<td>4.44 (.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>2.53 (.46)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4
Means (SE) and analysis of variance results for context integration measures.

<table>
<thead>
<tr>
<th>Condition</th>
<th>M (SE)</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>4.97 (.219)</td>
<td>8.03</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Repetition</td>
<td>4.41 (.222)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>3.70 (.229)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5
Means (SE) and analysis of variance results for recall length and quality measures.

<table>
<thead>
<tr>
<th>Condition</th>
<th>M (SE)</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>4.46 (.366)</td>
<td>7.76 (.71)</td>
<td>1.18 (.168)</td>
</tr>
<tr>
<td>Repetition</td>
<td>4.90 (.355)</td>
<td>8.29 (.66)</td>
<td>.309 (.845)</td>
</tr>
<tr>
<td>Control</td>
<td>4.12 (.370)</td>
<td>8.21 (.68)</td>
<td></td>
</tr>
</tbody>
</table>

Table 6
Mean percent (SE) and analysis of variance results for questions measures.

<table>
<thead>
<tr>
<th>Condition</th>
<th>M (SE)</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>24.23 (1.78)</td>
<td>1.054</td>
<td>.350</td>
</tr>
<tr>
<td>Repetition</td>
<td>27.40 (1.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>24.32 (1.80)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Production

For the Production Measure, the ANOVA showed significant differences across conditions. Post-hoc analyses (LSD) revealed that the control condition differed significantly from both the repetition (p < .001; d = .44) and interactive (p < .001; d = .70) conditions. In addition, the interactive condition scores were significantly higher than those for the repetition condition (p = .03; d = .26). The results are presented in Table 7. Mean scores represent number of target words students used for each condition.

Summary of results

We found that, as hypothesized, both instructional approaches showed significant advantage over control for students’ recognition of word meanings. Also as hypothesized, the interactive instruction better enabled students to integrate words into context and to produce words associated with a picture, relative to the repetition and control conditions. Repetition instruction also provided reliably better outcomes than control on these two measures. Contrary to our hypothesis, the listening comprehension measure showed no differences among conditions.

Discussion

The purpose of this study was to examine the effects on kindergarten children of two approaches to vocabulary instruction and a control condition using a progression of language processing tasks. We developed a series of measures to capture what students were able to do with the words they had been taught. The measures were designed to tap a progression from lower-order processing, represented by recognition of word meaning, to higher-order processing, represented by context integration, listening comprehension, and production. In this section, we discuss what the measures revealed about students’ understanding and use of words following instruction and consider some limitations and finally the implications of the study.

Meaning recognition

We initiated the progression of assessments with the Meaning Recognition Measure as an attempt to capture growth of word
knowledge. The finding that the control condition differed from the instructional conditions was not surprising. Although simply reading aloud to children likely enhances their vocabulary, these effects are minimal after only one passive encounter with words in a story context. The lack of difference between outcomes on the recognition measure for the repetition and interactive conditions was expected. It has been shown previously that instruction that provides practice on definitions is successful in affecting students’ word knowledge (Biemiller & Boote, 2006; McKeown et al., 1985).

**Context integration**

The slight advantage for the interactive condition may derive from the variety of encounters experienced in the instruction, which may have prompted students to build a richer set of connections to each word. For example, for the item asking why someone had to insist that his friend go on a Merry-go-round, richer connections might have allowed students to override the conventional reaction to a Merry-Go-Round and build a mental model of Merry-Go-Round riding that accommodated the potentially incongruent idea of insisting a child go on the ride. This interpretation is supported in that children in the interactive condition were more likely to provide responses such as: “Maybe he got sick to his stomach” and “It was scary—it went too fast.” Responses in the repetition and control conditions were more likely to suggest that students responded to the general context of going on a Merry-Go-Round, such as, “It was fun.”

**Listening comprehension**

The listening comprehension task represented text comprehension, which is a major eventual goal of vocabulary instruction. We assessed listening comprehension with recall and questions, largely based on success with such tasks in past research with third, fourth, and fifth graders (Beck et al., 1982; McKeown et al., 2009). However, many factors influence text comprehension beyond vocabulary knowledge, such as ability to sequence ideas, knowledge of story structure, knowledge of syntax, short-term memory limits, and referential understanding, that may overwhelm semantic processing. This may particularly be the case with students as young as kindergartners, who may be less likely to have these other aspects under control. Perhaps this is part of the reason that we found no differences among the conditions in the current study for our measures of listening comprehension.

The data from our study demonstrate that our participants found the recall task particularly difficult. Consider that the students produced sparse recalls, averaging about 14% of the text across all conditions. Contrast this with between 27% (McKeown et al., 1985) and 33% (McKeown et al., 1983) of a text containing newly taught words recalled for fourth graders in past vocabulary studies. Studies by Blank and Frank (1971) and Tucker (2001) also found that kindergartners produced very little language in response to recall prompts.

Students in our study also scored low on questions for the listening comprehension task. This result differed from Coyne et al.’s (2010) finding for their narrative comprehension task. Their findings of a positive trend in favor of the intervention group over a no treatment control might be traced to two distinct differences from ours, one a task difference and the other a treatment difference. The task difference is that in Coyne et al., the oral reading of the story narrative was accompanied by showing a wordless picture book, which represented the plot line of the story. The treatment difference was that Coyne et al.’s intervention was considerably longer. Each of our instructional cycles was seven days, whereas the treatment in Coyne et al.’s work was 18 weeks in length. Length of the intervention may make a difference, perhaps, as accumulation of experiences over the course of an intervention in making connections to new words, and extended practice with fitting new words into their language may provide general benefits to learners' comprehension of connected text. Some findings in the literature may seem to indicate that duration of intervention does not matter (Elleman et al., 2009; Marulis & Neuman, 2013; Mol et al., 2009). However, findings from these reviews do not speak to the effect of duration of vocabulary intervention on comprehension using taught words. The studies reviewed either focused on word knowledge and not comprehension (Marulis & Neuman, 2013) or mainly measured comprehension via standardized tests.

Thus, we see three possibilities for lack of differences on the listening comprehension task. One is that the students in the instructional conditions did not learn the words well enough to impact text comprehension performance. Although our measures of word knowledge, context integration, and production indicated that growth in understanding the words resulted from instruction, it may be that the knowledge did not reach sufficient quality or strength to affect comprehension. The second possibility is that a longer intervention is needed to yield text comprehension effects in order to influence general as well as word-specific processing. The third possibility is that the tasks of recalling and responding to questions about a text are too difficult for many students as young as kindergarten, possibly because the semantic aspects of the task are overshadowed by other requirements. Sorting out the three possibilities is not possible in the present case. Engaging in pilot work, as we did with the Content Integration task, might have allowed us to either discover that the task was a poor fit for kindergartners or be able to rule out its difficulty as the cause of poor performance.

**Production**

The production task explored students’ ability to use newly learned words in response to a picture that depicted features related to the words. Learning words well enough to express them is an important learning goal, and having words in one’s productive vocabulary is generally viewed as a good measure of word ownership. In addition, this task represents another process related to comprehension, ease of access to word meanings (Richter, Isberner, Naumann, & Neeb, 2013). If a reader or listener is to understand a word within a stream of language, he or she must be able to readily access a meaning for the word that fits the context. Being able to produce the word when seeing something that can be associated with the word suggests an ease of access that may make way for comprehension.

**Limitations**

We acknowledge a number of limitations to the study, some caused by necessary choices, others that could be resolved in future work, and still others that are inherent to work in the vocabulary field. One decision that could have affected outcomes and their magnitude was the length of the interventions. For each student, instruction in each condition lasted seven days. Our intent with this research was to focus on differences in instruction and try to pinpoint the differential effects. But the brevity of the interventions may have inhibited our ability to capture effects that may accumulate from ongoing attention to vocabulary in classroom instruction.

An issue in the design of the intervention was balancing encounters with each taught word across instructional conditions. We chose to do that by using the number of encounters—the number of activities in which a word was presented or discussed. Another method that has been used is “instructional intensity” (Baumann, 2009). Instructional intensity comprises time spent on each word. Yet prescribing amount of time to be spent in activities can be
difficult with instruction that is not rote. For example, some activities might produce longer discussion, or students might spend more time recalling a certain word’s definition or generating an associated word or definition.

Our choices of assessments created some limitations. First and foremost, we chose to use experimenter-designed measures in order to capture aspects of language processing that are not captured by extant measures. The psychometric properties of such measures may not be as clear, but their relevance to the outcomes sought is much greater than with standardized measures. Another measurement limitation was the lack of a delayed condition. Clearly, the ultimate goal of vocabulary instruction is for students to sustain the knowledge they gain. However, in making decisions about assessments, the amount of student and classroom time devoted to testing needs to be considered. We chose to present a range of assessments, which limited our flexibility to add a delayed testing condition.

Our use of story recall, as discussed earlier, limited the conclusions we could make about the outcomes. Because students’ recall was so low overall, we were not able to discern if the issues lay in the task itself or whether there was no effect on text comprehension. We could address this issue in future work in several ways. We could add a story with only common words, which would inform the difficulty of the task of recall itself; we could make the task more accessible by using a picture book to accompany it, as done by Coyne et al. (2010).

**Implications**

Implications from the present study relate to vocabulary instruction and vocabulary assessment. The consensus about effective vocabulary instruction, reached over recent decades, is that effective instruction for higher-order goals calls for both definitional and contextual information about taught words, multiple encounters with the words, and attention to active processing (Baker, Shanahan, Linan-Thompson, Collins, &Scarcella, 2007; Baumann, Kame’enui, & Ash, 2003; National Reading Panel, 2000; Stahl & Fairbanks, 1986). Our results support and extend this consensus.

Our results support the consensus in that differences on two measures of higher-order processing favored the cognitive processing-based instruction. The advantage for the interactive instruction on the context integration and production measures was small, which is not surprising as the differences in the instruction were small but targeted. For both the repetition and interactive conditions, activities were whole-class requiring oral responses and presented in game-like formats. Students received the same definitions and the same number of encounters. The key difference was that the interactive encounters featured a variety of contexts and involved students in interacting with the words. Although small, the differences are potentially meaningful because they are theoretically grounded, deriving from approaches that engage different processes. The repetition approach assumes that word learning can develop from repeated exposure to a single contextual instance combined with presentations of a generalized word meaning (i.e., a definition)—activities that recruit mainly memory and association processes. The interactive approach hypothesizes that word learning requires a learner’s actively abstracting information across varied word encounters, which recruits complex processes of inference and integration.

Our results extend the consensus about effective vocabulary instruction by demonstrating that instruction that prompts active processing allowed children as young as kindergarten to have more success in tasks that tapped higher-order language processing relative to instruction that offered repeated readings and word meaning practice or story reading only. Further, the results of both the repetition and interactive instruction demonstrate that young children can learn Tier-2 words, the kinds of words that are characteristic of written language, and can learn them to the extent that enhances their ability to comprehend contexts in which such words are used. These results in conjunction with the substantial evidence of gaps in very young children’s vocabulary at different SES levels and the pervasive effects of those gaps on children’s literacy development as children move through school (Snow et al., 1998) suggest that such early support is warranted. A beneficial next step might be to examine the extent to which explicit attention to vocabulary as ongoing classroom practice might be a resource in mitigating the gap.

Regarding vocabulary assessment, Pearson et al. (2007) have noted that this area of study is “undernourished.” A key aspect of the undernourishment is that studies of vocabulary frequently measure only meaning recognition-type outcomes that require only association with a definition. If studies provide multiple measures, additional measures are often standardized tests of general vocabulary or comprehension, which are not sensitive to short-term vocabulary growth, or passage-level comprehension tests (Ellem et al., 2009). Meaning recognition and text comprehension measures seem to represent two ends of a continuum of outcomes of vocabulary instruction, leaving a wide gap in our ability to understand the learning that took place. In order to learn more about the benefits that instruction may provide, we need intervening measures that require more than a definitional association but do not call for the complex set of processes needed for comprehending connected text.

By creating measures that tapped different points along a continuum of processing, we believe we have provided some of the “nourishment” for assessment that Pearson et al. (2007) call for. We see the context integration task as a particularly apt measure of semantic processing in that it taps what a reader needs to do when meeting a word in a novel context, at the point of occurrence of that process. The task calls on the student to make sense of the context based on the implications of the word’s meaning. We also may have provoked some assessment nourishment by identifying issues with using text comprehension measures with young children. Perhaps our use of the measure and the issues we identified will further nourish a more productive and informative use of such measures in future studies.

Ellem et al. (2009) and Pearson et al. (2007) serve to highlight challenges in the field’s search for resolving the relationship between vocabulary and comprehension. Pearson et al. posit that current measures of vocabulary may be inadequate to document the relationship between word learning and global comprehension. Ellem et al. echo the conclusion of Pearson et al., citing the inadequacy of measurement to fully capture the vocabulary/comprehension relationship as a major obstacle to identifying the qualities of effective vocabulary instruction. More targeted measures could allow more precise measurement of children’s learning, allowing us to develop more detailed and comprehensive understanding of how particular instructional activities or experiences support the kind of vocabulary growth that leads to comprehension enhancement.

**Author note**

We gratefully acknowledge the Spencer Foundation for its support of the research described in this article. The opinions expressed do not necessarily reflect those of the Foundation, and no official endorsement should be inferred. We also acknowledge with appreciation Ronette Blake for her data analysis work and her assistance with data collection and scoring, and Cheryl Sandora for her assistance with data collection and scoring.


