Word Knowledge and Comprehension Effects of an Academic Vocabulary Intervention for Middle School Students

Margaret G. McKeown
University of Pittsburgh
Amy C. Crosson
Pennsylvania State University
Debra W. Moore
Isabel L. Beck
University of Pittsburgh

This article presents findings from an intervention across sixth and seventh grades to teach academic words to middle school students. The goals included investigating a progression of outcomes from word knowledge to comprehension and investigating the processes students use in establishing word meaning. Participants in Year 1 were two sixth-grade reading teachers and 105 students (treatment n = 62; *control* n = 43) *and in Year 2, one seventh-grade reading teacher and 87 students (treatment* n = 44; *control* n = 43) *from the same public school. In both years, results favored instructed students in word knowledge, lexical access, and morphological awareness*

MARGARET G. MCKEOWN is a senior scientist at the Learning Research and Development Center and Clinical Professor at the School of Education, University of Pittsburgh, 3939 O'Hara St., Pittsburgh, PA 15260; e-mail: *mckeown@pitt.edu*. Dr. McKeown's work involves applying theory and cognitive research to practical, current problems of classroom teachers and their students. The focus of her work is reading comprehension and vocabulary.

AMY C. CROSSON is assistant professor at the Pennsylvania State University, University Park. Her research focuses on literacy interventions to improve word knowledge, comprehension, and argumentation of students in underresourced schools, including language minority students.

DEBRA W. MOORE is research assistant professor and co-director of the Pennsylvania Alternate System of Assessment (PASA) at the School of Education, University of Pittsburgh. Dr. Moore's research interests include applied issues in educational measurement, testing, and accountability systems, especially as applied to alternative assessments.

ISABEL L. BECK is professor emerita in the School of Education at the University of Pittsburgh. She has conducted research and published widely in the areas of decoding, vocabulary, and reading comprehension.

on researcher-designed measures. In Year 2, small advances were also found for comprehension. Transcripts of lessons shed light on processes of developing representations of unfamiliar words.

Introduction

A preponderance of evidence suggests that too few middle school students read at levels needed to learn from required school texts (American Institutes for Research, 2006; Heller & Greenleaf, 2007; National Center for Education Statistics, 2014). The most recent National Assessment of Educational Progress (NAEP) results show that only 36% of eighth graders read proficiently. Vocabulary demands of school texts are cited as a major contributor to reading problems. Harmon, Hedrick, and Wood (2005) assert that in the face of such demands, "students need to have a well-developed word knowledge base" to handle reading across content areas (p. 262). While comprehension is complex, involving the coordination of multiple cognitive processes and knowledge, the lynchpin of deep comprehension is *word knowledge* (Perfetti & Stafura, 2014). Word knowledge and reading comprehension are intimately linked (e.g., Adlof, Catts, & Little, 2006; Cromley & Azevedo, 2007; Tannenbaum, Torgesen, & Wagner, 2006) as abundant high-quality lexical representations characterize skilled readers relative to less skilled readers (Perfetti, 2007).

Attention to Vocabulary Instruction and Research

Despite the consensus on the key role of vocabulary knowledge, there is a lack of attention to vocabulary in schools, going back decades, from Durkin's (1978) work showing that 3% of instruction in reading dealt with word meanings, to Blanton and Moorman's (1990) work documenting that 6% of observed reading time focused on vocabulary. More recent studies include Connor et al.'s (2014) work with 27 third-grade classrooms, which found that teachers spent an average of five minutes daily on oral language and vocabulary instruction. Snow, Lawrence, and White's (2009) rationale for the Word Generation program suggests that this same dearth occurs at middle school.

When instruction does occur, it is rarely more than practice of definitions or synonyms. For example, Apthorp et al.'s (2012) vocabulary intervention study in kindergarten through fifth grade included an examination of transcripts from 26 business-as-usual classrooms. Half of the talk devoted to vocabulary was the teacher providing rote information to the students, usually a brief definition. The other half showed teachers calling on students, most often to request a definition or a synonym.

Vocabulary as a research topic has also experienced lack of attention, but unlike classroom issues, it has seen strong periods of resurgence, mostly at the elementary level. The field has produced relatively few studies focusing on middle and high school (cf. Baumann, Kame'enui, & Ash, 2003; Elleman, Lindo, Morphy, & Compton, 2009; Hairrell, Rupley, & Simmons, 2011). Vocabulary studies with adolescent learners have tended to focus on specific disciplines such as science (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009) or social studies (Vaughn et al., 2009), with two notable exceptions. Word Generation (Snow et al., 2009) and ALIAS (Lesaux, Kieffer, Faller, & Kelley, 2010; Lesaux, Kieffer, Kelley, & Harris, 2014) were designed to teach general academic words drawn from the Academic Word List (AWL; Coxhead, 2000) to middle school students. Word Generation is implemented in classrooms across disciplines, beginning in language arts class, then used across math, social studies, and science. Units of instruction are built around a short expository text on an issue of interest to students that uses the academic words to be taught. The instructional activities relate to that issue. ALIAS was also designed around a single expository text per two-week instructional unit and a strong focus on writing using the taught words. Instruction is designed for 45 minutes four times per week. Both programs have reported results that give the promise of effects on comprehension, but effects have been elusive (Lesaux et al., 2010, 2014; Snow et al., 2009).

Purpose

This report is based on the implementation of an instructional intervention, Robust Academic Vocabulary Encounters (RAVE), aimed to enhance academic vocabulary knowledge and comprehension across two years of middle school, sixth and seventh grades. Like Word Generation and ALIAS, RAVE teaches general academic words from the AWL. However, it is distinct from those programs in its focus on a range of expository contexts from authentic domains of use and its systematic and deliberate sampling of contexts to provide multiple senses and uses for each word. The activities in RAVE were designed to guide students to create multifaceted representations of word meanings based on elements of meaning similarity and difference demonstrated by various contexts.

Our purpose in reporting the work is to contribute to greater understanding of students' learning processes and outcomes from participating in the program, that is, what were students able to do with words they were learning and how did they process elements of instruction that led to those outcomes. To explore outcomes, we present results from a progression of measures that tap proximal to intermediate to distal outcomes, and to explore learning processes, we examine students' responses to lessons in which they were initially establishing word meaning from information provided during instruction.

Theoretical and Empirical Foundation

RAVE is designed to teach general academic words through instruction that provides multiple varied contexts, active processing of words and word uses, and attention to morphology through instruction in Latin roots. In this section, we describe the theoretical foundation that provides the rationale for those design decisions and how we instantiated that foundation into the lessons.

Theoretical Foundation

Historically, a strong link exists between vocabulary knowledge and reading comprehension, dating back to factor-analytic studies (Carroll, 1971; Davis, 1944; Singer, 1965; Spearritt, 1972). This link created the illusion that problems in comprehension could be addressed by promoting learning word meanings, mostly in the form of definitions (Beck & McKeown, 1991). That remedy proved mostly ineffective, and instructional studies in vocabulary beginning around the mid-1970s began to upgrade instruction to add components that researchers hypothesized might be needed to promote vocabulary knowledge that would contribute to comprehension. The field began to coalesce around a set of components, as indicated in review studies by Mezynski (1983) and Stahl and Fairbanks (1986). These were: contextual and definitional information, multiple encounters with words in various contexts, and active engagement with words. This consensus has held, as indicated by more recent work (National Reading Panel, 2000; Wright & Cervetti, 2017).

Over the past decade or so, theoretical perspectives have been fleshed out to address more precisely questions of the nature of word knowledge and how we learn words. Key work in the area includes Nagy and Scott's (2000) on vocabulary processes and Perfetti and Hart's (2002) on the lexical quality hypothesis. Nagy and Scott focus on the complexity of word knowledge, and they offer five aspects: incrementality, that knowledge of a word grows gradually, increasingly approximating mature understanding; polysemy, that word meanings are inherently flexible, with multiple senses and nuances; multidimensionality, that words have multiple forms and features, such as spoken, written, grammatical, register, collocational behavior, and associations with other words; interrelatedness, such that, for example, learning *cool* is easier if you know *hot, cold*, and *warm*; and heterogeneity, that essential knowledge about a specific word depends on the type of word it is.

Nagy and Scott (2000) make two important points about acquiring word knowledge given its complex nature. One is that only "exceptionally rich and multifaceted instruction" could provide all that is needed to be learned about each word (p. 273). The other is that what learners need to know about words is primarily procedural, how to use words, rather than knowing declarative meanings.

Perfetti introduced the lexical quality hypothesis to describe word knowledge (Perfetti, 2007; Perfetti & Hart, 2002). *Lexical quality* is the degree to which a learner's mental representation of a word specifies its phonological and orthographic form and its meaning both precisely and flexibly. The representation must be precise so that a word is quickly recognized when heard or read, yet all three components need to be flexible enough to accommodate variations, such as varied pronunciations required by different parts of speech or meaning variations depending on context.

Being able to quickly recognize a word embodies fluency of access to its meaning, or as termed by Reichle and Perfetti (2003), availability of the word's features, and it is essential for comprehension (Beck, Perfetti, & McKeown, 1982; Perfetti & Adlof, 2012; Richter, Isberner, Naumann, & Neeb, 2013). Fluent access to lexical representations releases working memory resources, which can then be directed toward higher-level comprehension processes.

Flexibility develops as learners build lexical representations into connected networks of associations to a word and across similar words based on multiple informative encounters (Reichle & Perfetti, 2003). A word's meaning becomes generalized across encounters so that it loses connection to specific contexts, allowing it to be applied flexibly to new contexts. Flexible knowledge allows a learner to integrate the meaning of a word into a novel context so that the learner can make sense of the context as a whole. The role of this integration process was noted by Yang, Perfetti, and Schmalhofer (2005, 2007), who asked participants to perform a variety of word learning and comprehension tasks while measuring brain responses through event-related potentials (ERPs). Skilled readers readily integrate a word with their understanding of a text while less skilled readers are slow and less successful at such integration (Perfetti & Stafura, 2014). These recent findings about integration make contact with older work that found that promoting learners' integration of word meaning into contexts enhanced comprehension effects (Jenkins, Pany, & Schreck, 1978; Kame'enui, Carnine, & Freschi, 1982).

A third perspective on word knowledge and word learning characterizes vocabulary learning as "the word-learning burden" (McKeown, Deane, Scott, Krovetz, & Lawless, 2017). This perspective emphasizes that encounters with words provide not only information on word meaning but also knowledge of generative patterns of syntax, semantics, and morphology. Syntactic patterns comprise how words fit into sentences, semantic patterns comprise how word meanings are extended through polysemy and other mechanisms to further describe situations and senses, and morphological patterns pertain to how words are constructed from root words, affixes, and bound roots. Beyond these patterns, knowledge of specific words also includes features such as register, collocational patterns, and semantic category.

Patterns of syntax and semantics can be gained through multiple encounters with words in context, and aspects of them are often directly

included in instruction about individual words as well. Recommendations to include morphology in vocabulary instruction are becoming more frequent. These recommendations stem from a strong body of research showing that knowledge of morphology contributes to reading comprehension (Anglin, 1993; Carlisle, 1995, 2000; Nagy, Berninger, Abbott, Vaughan, & Vermeulen, 2003). Evidence of instruction in morphology that leads to enhanced comprehension is not yet clear. Some studies suggest that the effects are equivocal or small (Baumann et al., 2003; Curtis, 2006; Goodwin & Ahn, 2013). Results of morphological instruction showed that students often learned the meanings for the word parts they were taught but rarely generalized that to the learning of new words (Bowers, Kirby, & Deacon, 2010; Curtis, 2006). However, other evidence has been more promising. Goodwin and Ahn (2013) conducted a meta-analysis of studies with students in the elementary and middle school, which found a statistically significant mean effect of morphology instruction for vocabulary (mean d = 0.34). Bowers et al.'s (2010) meta-analysis provided evidence of enhanced spelling and vocabulary learning across 21 morphological interventions. Although evidence of transfer to new words and reading comprehension was often small, two characteristics made transfer more likely: integration of the instruction into other aspects of literacy instruction and inclusion of a problem-solving stance.

The theoretical approaches of Nagy and Scott, Perfetti, and McKeown et al. differ in their terms and emphases, yet they all make clear that there is a great deal to be learned about words beyond a static description of a word's meaning. All three approaches also signal that although there is some systematicity to how we use and manipulate words, it is rather unpredictable and very word dependent as far as the information needed about a word to apply it for successful comprehension.

Word Selection

The type of words taught to students matters. The vocabulary burden of academic texts in middle and high school suggests the importance of providing students with direct vocabulary instruction that is efficient and effective, which inherently encompasses selecting words to teach that will be optimally useful. As a perspective on how words can be categorized for their usefulness, consider the four categories that Nation (2001) describes: high frequency, academic, technical, and low frequency. Academic words are words common across domains of academic texts, as exemplified by Coxhead's (2000) AWL. Nation's technical words are those related to a specific topic or subject area but not common beyond it. Nation's academic category roughly corresponds to Tier 2 words, which are described as general words that appear across domains and that characterize text rather than conversational language (Beck & McKeown, 1985; Beck, McKeown, & Kucan, 2002, 2013). However, Tier 2 also includes words from literature. Although

there is much overlap between literary and academic texts, the two corpora are distinct (Gardner, 2004; Hiebert & Cervetti, 2012). Many literary words tend to be stronger and more imageable, for example, *remorse, solace*, and *surreptitious*, than words that typify academic discourse, such as *coordinate, variable*, and *duration*.

Our intent was to present academic words that go across disciplines and provide core meanings of the words that support more general uses rather than to focus on disciplinary applications of academic words. We chose Coxhead's (2000) AWL as a useful set of words for instruction as those words have empirical backing as occurring frequently in academic texts across disciplines. Although the AWL was created from university textbooks, the words do appear in texts for younger students. For example, in a corpus of 385,414 running words developed from published materials for fifth and sixth grades, 485 of the 570 AWL words appear (Gardner, 2004). All words taught in RAVE were selected from the AWL.

Instructional Design

Research conducted across decades and grade levels has supported the consensus that vocabulary instruction that is likely to support comprehension is based on multiple encounters with words in varied contexts and encouraging active processing (Beck et al., 1982; Bos & Anders, 1990, 1992; Carlo et al., 2004; Coyne et al., 2010; Margosein, Pascarella, & Pflaum, 1982; McKeown, Beck, Omanson, & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985). But because learners need a wide variety of information about words to become successful comprehenders, perhaps we can identify more focused experiences around multiple exposures and active engagement that more precisely embody theoretical aspects that are key to achieving effective word knowledge. Toward that end, we designed RAVE instructional activities to focus on five key aspects derived from our theoretical foundation: (a) polysemy around a core meaning, (b) enriching semantic representations through exposure to patterns of use, (c) building fluency of access to word meanings, (d) direct focus on integration of word meaning and context, and (e) morphology. In this section, we describe how we instantiate those features in the program. In a later section, we review the program format and sequence of lessons.

All of these aspects overlap in both theory and practice. For example, the ability to integrate word meaning and context is supported by understanding polysemous senses or nuances of word meaning and by familiarity with generative patterns of word use that can be derived from exposure to multiple contexts. As such, activities in RAVE emphasize polysemy, exposure to patterns, fluency, integration, and morphological analysis in various ways, with emphases shifting among the elements and many activities linking to several features.

The introductory activity in RAVE targets polysemy, pattern exposure, and integration of word and meaning. Introductions begin with two authentic contexts that illustrate how a word is used in various domains. This is important because the contexts containing academic words are often beyond everyday familiar experiences, and contexts from various domains embody senses or nuances that may vary enough to disrupt comprehension. Consider the AWL word *expose*, which can appear in contexts about substances such as radiation and chemicals as well as contexts about art and culture.

Another design element that addresses polysemy is the "friendly definition" in the introductory lesson for each word, which is built around a core meaning that unites various senses of the word. For instance, the senses of *expose* mentioned previously seem disparate, yet they share a core meaning of, roughly, "allowing something to be shown or revealed." Specifically, RAVE's friendly definition is: "If you expose something, you let it show or make it known."

Integrating meaning and context is directly addressed in the introductory lesson when students are asked two questions about how the meaning of the word fits the contexts. The initial context for *expose* describes blasting off the tops of mountains to expose layers of coal. The goal for the first integration question is to have students say something like "When mountain tops are blasted off, you can see the coal underneath." The second context for *expose* is about exposing students to jazz music, after which students are asked what the two contexts show about how the word can be used, targeting both integration and polysemy. For that question, we expect a response such as "You can expose something that is hidden, like coal underground, or you can expose people to new ideas by showing them new things."

Subsequent lessons provide encounters within contexts that typify a word's use. We selected typical uses of words based initially on our own experiences as language users, but we also used WordNet (Fellbaum, 1998) and the Corpus of Contemporary American English (COCA; Davies, 2008) as resources to assure coverage. For example, encounters with *expose* include being exposed to information and experiences (middle school students exposed to what high school will be like, exposing students to the history of the Civil War), exposing hidden identity (expose someone's mistreatment of his dog, expose people using counterfeit money, expose someone's fake identity), and physically exposing something not visible (exposing brick underneath paint).

Activities were developed around the various contexts to further reflect the key theoretical aspects. For example, RAVE included activities to enrich representations by building awareness of semantic features that discriminate words, also prompting word and context integration. In one such activity, students were asked to explain what the school art teacher would do to "*expose* her students to watercolor painting" versus "*refine* her students' watercolor painting skills." Students needed to examine the features of

each target word, choose the elements of each that apply to the scenario, and develop an explanation of how that word functioned in the context, such as "exposing students to watercolor painting might include showing examples of watercolors, while refining students' watercolor painting skills would include having them practice painting themselves." An activity aimed at supporting fluent access presented a range of situations and asked students to recognize as rapidly as possible which exemplified each of the target words. The situation for *expose* was: "Telling the truth about a group of people using counterfeit money."

Morphology was instantiated in RAVE using bound roots derived from Latin, such as *fin* in *finite*. Although English has a grammar and core vocabulary inherited from German, a significant portion of English vocabulary is Latinate (Baugh & Cable, 1978; Myers, 1966). In particular, 75% of the words on the AWL (Coxhead, 2000) are Latinate (Lubliner & Hiebert, 2011); thus, knowledge of Latin roots might benefit reading of academic texts. This approach is distinct from previous work in morphology instruction, which has focused mainly on derivational affixes, morphemes added to words to alter their form or meaning, such as *un-* and *-ed* (Goodwin & Ahn, 2013). Other types of morphemes in English include inflectional morphemes (e.g., suffixes that change tense or number) and morphemes that provide the semantic foundation of words, which can be either free-standing "root words" such as *school, run*, or *happy*, or bound roots. We refer to roots that carry a word's main semantic information as *lexical morphemes* because *lexical* relates to "the vocabulary of a language as distinguished from its grammar and construction" (Merriam-Webster.com, 2011).

An activity on morphology was included in the final lesson of the unit. Morphology activities were chiefly designed to develop students' awareness that words can be analyzed for their parts (i.e., lexical morphemes) and the resulting information can assist in understanding unfamiliar words and relationships among words. Fluent access to word meaning was addressed as instruction was designed to build fluency through familiarity with morphological structures.

The lessons also introduced root-related words—words that shared a root with the target word. Thus, for the word *consequences*, students worked with the Latin root *sequ*, meaning "follow," and the words *sequel* and *sequence*. Lessons included activities that prompted students to figure out how a root fit the meaning of a word or how an image demonstrated the meaning of a root-related word. Thus, lessons incorporated the recommendations of Bowers et al. (2010) that morphology instruction is more likely effective if it is integrated with other aspects of instruction and introduces a problem-solving perspective. (For a full discussion of this aspect of the project, see Crosson & McKeown, 2016.)

Method

The RAVE materials were designed for sixth and seventh grades, providing instruction for a total of 195 words, 99 in sixth grade and 96 in seventh grade. (For a list of all words taught, see the Supplementary Appendix in the online version of the journal.) We assessed a progression of proximal to distal outcomes in sixth grade in Year 1 and seventh grade in Year 2. In addition, in Year 2, outcomes on the distal measures were examined to compare students who had experienced RAVE in both years with students who had not experienced RAVE in either year.

Rationale for Methodological Approach

Progression of Outcomes

Assessment of vocabulary instruction outcomes has been quite restricted (Pearson, Hiebert, & Kamil, 2007), with most studies relying on measures that involve knowing definitions for the taught words. Although a major goal of vocabulary instruction is to enhance comprehension, measuring effects on comprehension is often neglected. Consider that two recent analyses of vocabulary instruction research did not include comprehension results for the studies they reviewed (Hairrell et al., 2011; Marulis & Neuman, 2013). When comprehension is addressed, it is most often with tandardized measures, which are not sensitive to short-term growth and thus rarely show positive outcomes of vocabulary instruction (Elleman et al., 2009; Pearson et al., 2007).

Even if gains are realized on standardized measures, such results reveal little about how students' facility with vocabulary was affected (e.g., Curtis & Longo, 2001; Nelson & Stage, 2007). Word knowledge tasks and standardized comprehension tasks represent extreme ends of a continuum from proximal to distal outcomes, leaving a wide gap in understanding what students can do with the words they have been taught.

Our study employed a progression of measures that tap a range of language processes involved in understanding and applying word meaning. We included a proximal measure of word knowledge, intermediate measures of comprehension of text containing instructed words and lexical access of instructed words, and distal measures to tap processes of morphological awareness and general comprehension. The word knowledge measure is considered proximal as it most closely aligns with what was directly taught. Testing comprehension and lexical access was considered intermediate as the measures focused on more general processes, and although they contained instructed words, they did not directly measure knowledge of those words. The morphological awareness and general comprehension measures were considered distal as both tapped more general processes and did not include instructed words. The focus on a progression of outcomes

emphasizes the range of knowledge needed to advance the field's understanding of vocabulary effects.

Vocabulary Learning Processes

Exploring the kinds of thinking that students engage in as they work to understand new words from instructional interactions may contribute to understanding the complexities and challenges of establishing word meaning. This understanding can in turn lead to more effective ways to support students' learning. In this study, we analyzed lesson transcripts to examine how students process instructional information and obstacles that impede processing. Much has been learned about comprehension processing by examining how students respond to text and various types of questions asked during reading (see Applebee, Langer, Nystrand, & Gamoran, 2003; Nystrand, 1997; Palincsar & Brown, 1984). There may be similar value in examining interactions during vocabulary instruction.

Research Questions

The impact of RAVE on student learning of academic vocabulary and comprehension was investigated using a quasi-experimental design with sixth graders in Year 1 and seventh graders in Year 2. The study addressed the following questions:

Research Question 1: Does RAVE instruction enhance students' knowledge of academic words?

Research Question 2: Does RAVE instruction enhance students' efficiency of access to word meanings for instructed words?

Research Question 3: Does RAVE instruction enhance students' comprehension of text containing instructed words?

Research Question 4: Does RAVE instruction, which includes lessons on morphology, enhance students' morphological problem-solving ability?

Research Question 5: Does RAVE instruction enhance students' general comprehension as measured by a standardized reading comprehension test?

Research Question 6: What do interactions during RAVE instruction demonstrate about the process of establishing word meaning?

Intervention Design

RAVE was designed around cycles of seven daily scripted lessons over about 22 weeks of the school year in both years of the study. The major design features of RAVE were:

- introductions to each word based on:
  - a pair of authentic contexts that displayed multiple senses or uses

    ○ a definition built around core meaning
    ○ prompts to integrate word meaning and context

- follow-up activities to promote active use of the words
- attention to morphology through Latin roots.

All lesson activities embodied scripted instructions for teachers to fol-low. Contexts were drawn from diverse print and internet-based resources and selected to display typical uses of the words as found in various domains. Contexts ranged from 40 to 90 words.

Materials included a teacher notebook of all lessons and individual stu-dent notebooks. The teacher's notebook contained all activities, and lessons were scripted with suggested language and the goals of the activities. Possible responses were also included. For example, introductory lessons (here, for the word *features*) began:

- T or S reads Context #1 aloud.
- To understand Context #1, ask:
    ○ **What do we know about carnivorous plants?**

<u>Ideas to develop</u>: They are unusual and beautiful and eat meat.

Student notebooks included copies of the contexts and definitions used in the introductory lessons and activity pages for most of the activities. Some of these were text for students to read or follow along, and some were pages that called for students' responses. In general, the program was conducted whole class, with some activities formatted for pairs or teams of students, but that was teacher's choice to do so. Instructional interactions were intended to promote discussion and sharing ideas. Teachers were encour-aged to ask students to explain their responses and get responses from mul-tiple students. Both teacher and student notebooks included a glossary of all target words.

Words were introduced across the first three lessons of the cycle, three words per lesson. Word introductions concluded with a quick activity to prompt students to use or make decisions about the word. For example, for the word *features*:

- Which *features* do you think are important to have on a cell phone? Why?

Lessons 2 and 3 began with a quick review of the words introduced in the previous lesson.

Over the remainder of the seven-day cycle, activities were provided to prompt students to interact with the words' meanings and uses. Activities were sequenced from ones that required simpler processing, such as

choosing a context that fits a word, to more complex activities, such as asking about relationships between words or using the words in writing. In general, Lesson 4 activities addressed each word by itself and were mainly oral response. In Lessons 5 and 6, students were asked to go across words, such as selecting which word fit a context or describing the differences between how two words would affect a context. In Lesson 6, a writing activity was included, always followed by students' sharing what they had written. Lesson 7 featured a brief assessment in the form of either matching word to definition or fill-in-the-blank sentences and the lesson on morphology, called "Becoming Aware of Language" (BAoL).

The BAoL lessons taught 18 roots in sixth grade and 16 different roots in seventh grade. Between two and four roots were taught per lesson, selected from the words taught in that cycle. Several root-related words, which shared a root with a target word from the lesson, were also presented in each lesson. These words appeared in only one lesson. A total of 95 root-related words were introduced across the lessons.

Year 1

Participants

The participants were two sixth-grade teachers and 105 students (RAVE $n = 62$; control $n = 43$) from all five sixth-grade classes in a public middle school within a working-class community in the northeastern United States. About 25% of the students were African American and the rest European American, and 55% received free or reduced-priced lunch. All spoke English as their first language. In reading, 59% of sixth- and seventh-grade students scored at proficient levels on the state assessment in 2012. Classes were heterogeneously grouped, and RAVE classes were chosen at random for each teacher. There was no difference in reading comprehension between the RAVE and control groups as measured by the Gates-MacGinitie Reading Test (GMRT; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2007) (Level 6, Form S) at pretest, $F(1, 102) = .008$, $p = .928$, $\eta_p^2 = .001$. The two teachers taught language arts and social studies for all students in sixth grade. They were European American females, each with about 15 years of teaching experience. Each teacher taught one control class and either one or two RAVE classes.

Materials

The RAVE materials for sixth grade provided instruction on 99 academic words over about 22 weeks, in 11 seven-day cycles interspersed with four review cycles. Each cycle taught nine words organized into themes intended to capture an element that the words shared (e.g., the cycle "Give or Take" includes *obtain, submit, compensate, compile, extract, restrict, prohibit,*

allocate, and *consume*). Lessons were from 10 to 20 minutes in length, depending on the activities. Students in control classes received a nearly identical dosage of vocabulary instruction that was a component of their basal reading series, *Storytown* (Beck et al., 2009).

Prior to implementation of RAVE, the research team provided the two teachers with a half-day workshop that reviewed research in vocabulary development and introduced the RAVE materials. Teachers received notebooks containing all materials for the lessons and detailed instructions about presenting each lesson. The teachers implemented the lessons daily as part of their reading/language arts block, which was 90 minutes long. The rest of the reading/language arts instruction was based chiefly on a basal reading program (Beck et al., 2009), although teachers also occasionally used novels and introduced their own writing prompts. The basal contained a daily vocabulary component, which was implemented only in the control classrooms.

Members of the research team observed lessons one to three times per cycle and met with the teachers about once per cycle to share feedback. Observations were informal, with observers keeping detailed field notes that were then shared among the research team. A systematic observation protocol was also developed and used twice per teacher to collect fidelity data. All lessons were audio recorded and transcribed.

Measures

Our goal of measuring a progression of outcomes drove our selection of measures. Three researcher-designed measures—a test of word knowledge for all taught words, a lexical decision task, and a passage comprehension task—were used to measure effects of learning the target words. We developed and administered an assessment of morphological awareness to assess the effects of instruction in Latin roots. We also included a standardized comprehension task. To address our goal of investigating students' processes of establishing word meaning, we analyzed transcripts of classroom interactions.

Word Knowledge Test

The Evaluation of Academic Vocabulary (EAV) was developed as a pre- and posttest of knowledge of all 99 instructed words. Each test item presented a word and four fill-in-the-blank sentences, and for each sentence, students had to decide whether the target word fit. The task was designed to capture depth of knowledge by assessing multiple aspects of word knowledge and use. The EAV most directly reflects the key aspects of polysemy and exposure to patterns of use. The test tapped polysemy as two of the sentences for a word could be correct, such as physical and mental senses of *confine*: "He will _____ the toddlers to the little yard" and "I had lots to say, but I had to _____my comments to the topic." The test tapped whether

students understood patterns of use well enough to reject foils that were inconsistent with those patterns. Foil types included an orthographic foil (*confine-combine*: "She wanted to _____all the ingredients into one recipe.") that could be easily dismissed, even with a low level of word knowledge, and a more difficult semantic foil with a prototypical association to the target word (*confine-jail*: "Prisoners often _____letters to keep in touch with their families."). For words for which we tested only one sense, we also included an unrelated foil, which contained no association with the target word but any word that could plausibly fit the sentence was the same part of speech as the target word. For example, "All the bananas had to be thrown away because they were _____." for the word *significant*.

Each item was scored 0 to 4, depending on the number of sentences that students correctly chose as matching or not matching the target word. The measure exhibited strong internal consistency (Cronbach's α = .91), and bivariate correlations with the reading vocabulary subtest from the GMRT ($r = .76$, $p < .001$) provided evidence of convergent validity.

Lexical Decision Task

A lexical decision task was administered following the intervention to assess fluency of access to target word meanings. Performance on this task reflects instruction by calling on rich lexical representations built from exposure to multiple contexts and connections between orthographic constituents and their meaning representations built from the morphology instruction.

In this computer-based reaction-time measure, students made decisions as to whether strings of letters shown on the screen were a word. Three sets of 20 words/nonwords were shown in random order using E-Prime (2002) software: words taught in RAVE, words matched to the RAVE words for frequency and orthographic complexity, and nonwords equated for orthographic complexity. The RAVE words comprised 13 words with the highest scores at pretest for all students plus 7 words that appeared in the school environment, such as science and social studies texts (e.g., *dominant* and *feature*). Our hypothesis was that because these words were already somewhat familiar to students, assessing access to them after RAVE instruction would provide a good test of whether experiencing RAVE improved fluency of semantic access, which plays a role in enhanced comprehension (Richter et al., 2013).

Text Comprehension Measure

A passage comprehension recall task was based on a text containing instructed words and a comparison text with untaught words from the AWL. This measure most directly reflects the key aspect of integrating words within contexts.

Such tasks used in prior vocabulary work were found to differentiate students who had learned target words from a control group (Beck et al., 1982) and also differentiate different types of instruction (McKeown et al., 1985). Two expository texts were developed and equated for text structure, number of words (325), number of AWL words, number of multisyllabic words, and readability. Students were asked to read the texts and provide oral recalls in individual sessions.

Recalls were scored on the number of clausal-length units recalled, weighted by importance such that main units were given 3 points, support units 2 points, and details 1 point. The texts each comprised 38 units: 5 main, 14 support, and 19 detail. Recalls were scored by one researcher, and 20% were then scored for interrater reliability. Exact agreement was 92%.

Morphological Awareness Measure

We developed a dynamic assessment to measure morphological analysis ability, which reflects instruction in morphological analysis around Latin roots. The measure was administered as a posttest. Dynamic assessment prompts students to answer a set of questions in a problem-solving sequence. After each step in the sequence, information from that step is provided so that students can proceed despite lack of knowledge of any specific step. As such, dynamic assessments provide rich information about students' thinking processes, especially for skills and knowledge that are still forming (Burton & Watkins, 2007; Poehner & Lantolf, 2010). Thus, this task assessed the process of applying morphological information rather than just the knowledge needed to do so. Similar measures have been used to investigate morphological problem solving of unfamiliar words by Anglin (1993) and Pacheco and Goodwin (2013).

The task consisted of seven sentences, each with a novel word containing a taught root. For example, "Most of their conversations were about the minutiae of daily life," with *minutiae* sharing the root *min* with the RAVE word *diminish*. The task was conducted with six students from each class, sampled to represent two each from Quartiles 1 to 3 on the GMRT pretest. Because of the prolonged absence of one student, we tested 29 students (17 RAVE and 12 control students). An independent samples *t* test confirmed no significant difference in prior reading achievement between the two groups ($t = .351$, $p = .363$). Students were shown each sentence and asked to explain it. Depending on a student's response, follow-up questions asked the student to identify the root (referred to as "word part" for the control students) and its meaning. This information was supplied if the student did not know it before the final question, which again asked the students to explain the sentence.

Responses were coded for (1) *recognition* of the root (0 = no, 1 = yes), (2) *meaning* of the root (0–2 depending on the amount of prompting), and (3) how successfully the meaning of the root was used to gain

comprehension of the context (0 = no use of root meaning, 1 = use of root meaning but applied inaccurately, 2 = use of root meaning to accurately interpret the context). Transcripts were blinded and coded by one member of the research team. Another researcher independently coded 20% of the transcripts, yielding 88% exact agreement.

General Comprehension Measure

A standardized test of reading comprehension (GMRT, Level 6 Forms S&T; MacGinitie et al., 2007) was administered pre and post. We acknowledge that it is unrealistic to expect greater gains from the intervention on a general standardized comprehension measure, but general comprehension is the ultimate goal, and there are few choices for such a measure (Pearson et al., 2007). Although gains in general comprehension may be too distal to expect, the RAVE instruction aligns with the consensus about the kind of instruction most likely to lead to comprehension gains. Further, RAVE offered specific practice in integrating word meaning and context, a key feature of skilled comprehension.

Fidelity of Implementation

Fidelity was measured with a transcript analysis of full seven-day instructional cycles from all RAVE classes in the fall and spring. In addition, classroom observations to capture instructional quality were conducted in the fall and spring in intervention and control classes.

Transcript analysis. Following McKeown, Beck, and Blake (2009), we analyzed transcripts of a sample of lessons to examine whether the teachers implemented the lessons as designed and assure a constant level of implementation across teachers and classes. Two members of the research team coded transcripts of all seven lessons in two RAVE cycles for all three classes, one from the fall and one from the spring. Transcripts were coded for the percentage of scripted questions and explanations in the teacher materials that were implemented.

Observations. Two members of the research team observed all RAVE and control classes once in the fall and again in the spring. The purpose of these observations was to collect contextual information and establish comparability of classrooms along broad dimensions not typically captured in transcripts but that may influence quality of implementation, such as student participation and classroom climate. All classes were rated on seven dimensions on a 3-point scale, and descriptors and examples were specified for each scale point. Aligned with previous observation tools (Lawrence, Crosson, Paré-Blagoev, & Snow, 2015; Matsumura, Garnier, & Spybrook, 2013), participation was measured as the percentage of students who

participated verbally in the lesson (1 = half or fewer students contributed verbally during the lesson, 3 = at least three-fourths contributed). Items to measure classroom climate were aligned with previous observation tools designed to measure teacher affect and climate (Lawrence et al., 2015; Matsumura, Slater, & Crosson, 2008), such as whether classroom climate is positive (1 = classroom climate was negative, i.e., students teased or insulted each other, lots of side talk; 3 = students treated each other with respect; when there was a need to focus students' attention, the teacher was able to do so quickly and respectfully). The same RAVE lesson in each of the three intervention classes and the same control lesson in the two control classes were observed on the same day. Reliability was calculated as percent exact agreement. Overall exact agreement (for each dimension at every observation) between observers was 86% in the fall and 91% in the spring.

Administration of Measures

Pretesting (EAV and GMRT) was conducted in the week before instruction began in the fall. Posttesting, which included all measures, was conducted in the two weeks following completion of instruction. The EAV, GMRT, and lexical decision tasks were administered to all students in each classroom during their language arts period. The passage comprehension task was administered individually to each student, and the morphological awareness task was administered individually to the 29 participating students.

Results

Analysis

All analyses were conducted in SAS (the PROC MIXED or PROC GLM procedure, controlling for the Grade 6 GMRT pretest score).[1] Comparisons were performed using the LSMEANS command with a Tukey adjustment. The assumption of sphericity was met in all cases, and the assumption of normality was met except where noted. Due to the flexibility in modeling available and complexity of the estimation used with the PROC MIXED procedure, effect sizes could not be calculated (Selya, Rose, Dierker, Hedeker, & Mermelstein, 2012; Tippey & Longnecker, 2016). However, effect sizes were calculated when using the PROC GLM procedure.

Word Knowledge Test

A 2 × 2 mixed analysis was performed on EAV test scores as a function of treatment and time controlling for GMRT pretest. There were two groups, RAVE ($n = 62$) and control ($n = 43$). There was a significant interaction of EAV scores between treatment and time, $F(1, 102) = 55.22$, $p < .001$. There was also a significant main effect of group, $F(1, 102) = 13.58$, $p < .001$, and time, $F(1, 102) = 125.59$, $p < .0001$. While there was no significant

Means and Standard Deviations for Sixth-Grade Evaluation of Academic Vocabulary Word Knowledge Test Scores as a Function of Treatment

	n	Pretest	*SD*	Posttest	*SD*
Robust Academic Vocabulary Encounters (RAVE)	62	240.00	29.04	283.98	41.96
Control	43	244.42	38.07	253.33	44.32

difference between pretest scores between groups, there was a significant difference between posttest scores between groups in favor of the RAVE group, $t(102) = 7.04$, $p < .0001$. Additionally, posttest scores were significantly higher than pretest scores for the both the control and RAVE groups, $t(102) = 2.87$, $p = .03$ and $t(102) = 12.36$, $p < .0001$, respectively (see Table 1).

Lexical Decision Task

A mixed analysis was performed on the reaction times with word type (RAVE vs. matched) as the within-subjects variable and group (RAVE vs. control) as the between-subjects variable controlling for GMRT pretest. Only response times for correct decisions were included in the analysis. (Accuracy rate was 91.11%.) There was a main effect of word type such that RAVE words were responded to faster than matched words, $F(1, 97) = 11.68$, $p < .001$, but there was no main effect of group, $F(1,9\ 7) = 3.04$, $p = .08$. Importantly, there was a significant interaction between group and word type, $F(1, 97) = 7.35$, $p = .008$. Students in the RAVE group had significantly faster reaction times for RAVE compared to matched words, $t(97) = 5.06$, $p < .0001$, while the control group did not, $t(97) = .44$, $p = .97$ (see Figure 1). Also, the RAVE group had significantly faster average reaction time on RAVE words compared to the control group, $t(97) = 2.75$, $p = .04$.

Text Comprehension Measure

A mixed analysis was performed on the recall scores from the two texts controlling GMRT pretest. The within-subjects variable was text with two levels (RAVE words and untaught AWL words). The between-subjects variable was group (RAVE and control). The analysis indicated no significant interaction between group and text and no significant main effect for text type or for treatment (see Table 2).

Morphological Awareness Measure

Separate generalized linear models (GLMs) were run to analyze the performance of RAVE and control students on the recognition, meaning, and
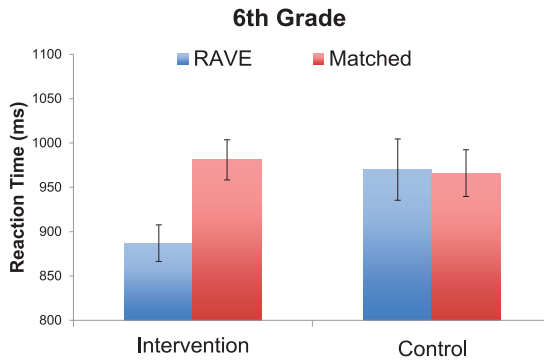
Figure 1. **Mean reaction time in milliseconds to Robust Academic Vocabulary Encounters (RAVE) and matched words for intervention and control students on lexical decision task in sixth grade.**

Table 2
Means and Standard Deviations for Sixth-Grade Text Comprehension Scores as a Function of Treatment

Text	n	RAVE		n	Control	
		M	SD		M	SD
Pets	62	14.30	7.70	43	13.36	6.74
Circus	62	14.69	5.38	43	13.80	7.39

Note. The pets passage contained untaught words from the Academic Word List (Coxhead, 2000); the circus passage contained RAVE words. RAVE = Robust Academic Vocabulary Encounters.

comprehension dimensions. After controlling for GMRT pretest, RAVE students scored significantly higher than the control group on the recognition, $F(1, 113.43) = 41.41$, $p < .001$, partial $\eta^2 = 0.61$; meaning, $F(1, 28.61) = 6.02$, $p = .02$, partial $\eta^2 = 0.19$; and comprehension dimensions, $F(1, 39.82) = 7.80$, $p = .01$, partial $\eta^2 = 0.23$, suggesting small (for meaning and comprehension) to large (for recognition) effect sizes for these dimensions in this small sample (see Table 3).

General Comprehension Measure

A mixed analysis was performed on GMRT extended scale scores (ESS) as a function of time and treatment. The assumption of normality was met for the control group for pretest scores and the RAVE group for posttest scores.

Table 3
Means and Standard Deviations for Each Dimension of the Morphological Analysis Task Dynamic Assessment for Sixth Grade

Dimension		RAVE			Control	
	n	M	SD	n	M	SD
Recognition	17	6.53	1.66	12	2.50	1.57
Meaning	17	4.24	2.31	12	2.25	1.91
Comprehension	17	13.82	2.07	12	11.42	2.43

Note. Recognition possible maximum = 7; meaning possible maximum = 14; comprehension possible maximum = 14. RAVE = Robust Academic Vocabulary Encounters.

There was no significant interaction of time and treatment condition on GMRT ESS scores, $F(1, 102) = 0.84$, $p = .36$. There was no main effect of group on GMRT ESS scores, $F(1, 102) = 0.27$, $p = .60$. However, there was a main effect of time on GMRT ESS, $F(1, 102) = 16.84$, $p < .0001$. Simple comparisons between timepoints by group showed that posttest scores were significantly higher than pretest scores for both the control and RAVE groups, $t(102) = 3.28$, $p = .008$ and $t(102) = 2.48$, $p = .04$, respectively.

Fidelity of Implementation

For the transcript analysis, the lesson material was implemented 95% of the time across the three RAVE classes, indicating a high degree of fidelity, which is not surprising given the scripted nature of the materials. For observations, scores revealed high levels of participation and positive classroom climate across both RAVE and control classes. Scores are presented as the average for each condition in Table 4. The consistency between teachers and across conditions is not surprising as they collaborated closely and were experienced teachers who had developed respectful and efficient classroom routines.

Discussion of Year 1

Results from implementation in sixth grade provide preliminary evidence that RAVE is effective for improving students' knowledge of academic words and enhancing lexical access to instructed words. The EAV measure is a proximal measure, yet by attending to multiple senses and partial knowledge, it goes beyond traditional word knowledge measures to tap multidimensional word knowledge in context. Thus, it may be a more stringent measure of vocabulary growth. Improved lexical access suggests that the RAVE words may be processed more efficiently, thus allowing more mental resources to be available for comprehension (Richter et al., 2013).

Table 4
RAVE Observation Protocol Scoring Guide to Rate Quality of Instruction in Intervention and Control Classrooms in Sixth and Seventh Grades

Item	Sixth Grade		Seventh Grade	
	RAVE	Control	RAVE	Control
Teacher made an effort to encourage student participation in the lesson.	2.5	2	2	2
There was broad student participation in the lesson.	2.5	2.75	2.33	1
Students were highly engaged during the lesson.	2.5	2.5	2	2
Teacher's affect and demeanor communicated enthusiasm about the lesson and engagement with the content.	2.83	2.5	2	2
Teacher has established a positive classroom climate, enabling students to focus on lesson content.	2.83	3	3	3
Teacher was prepared for the lesson.	2.83	3	3	3
Timing and pacing of lesson maximized student learning.	2.33	2.25	2	3

Note. Scores are averaged across raters and across observations within each year. Sixth grade $n = 3$ RAVE and 2 control classes; seventh grade $n = 2$ RAVE and 2 control classes. RAVE = Robust Academic Vocabulary Encounters.

Results from the morphological analysis task demonstrate that RAVE students were more successful at applying meanings of roots to infer meanings of unfamiliar words to comprehend sentences than were control students, suggesting that RAVE students had developed some awareness of morphological decomposition as an analytic tool.

As to why no differences were demonstrated on the text comprehension (recall) task, there are several possible explanations. One is that perhaps the students did not learn the RAVE words sufficiently for them to play a role in comprehension or that what students knew about the words was not key to their comprehending the texts. A second explanation may reside with the use of expository texts rather than narrative texts, which had been used in similar tasks in the past. For example, expository text structure, relative to more familiar narrative text structure, may not have provided the support that allowed students' vocabulary knowledge to bolster comprehension. A third possibility is that the amount of recall students provided was too limited to demonstrate differences. The recalls averaged 8 text units, which is 21% of the text, which contrasts with between 27% (McKeown et al., 1985) and 33% (McKeown et al., 1983) of a text recalled for fourth graders

in past vocabulary studies. The length of the recalls may have been related to students' ability to respond to expository versus narrative text.

No differences were found for the standardized assessment of reading. Given the history of effects on standardized scores, this was expected (see Elleman et al., 2009). Gains on standardized assessments following interventions are rare for several reasons, chief among them that generalized improvement in comprehension from a vocabulary intervention may not be immediately measurable; we would expect more proximal results initially, such as comprehension of a passage containing taught words.

Year 2

Context of Year 2 Study

Year 2 occurred in the same public school as Year 1. Our original intent was to provide two years of the intervention to one group of students with another group serving as comparisons. Despite our planning with the school district to do so, the district's evolving parameters for assigning students to classrooms prevented us from fully carrying out that plan. Thus, in both RAVE and control classrooms, we had a mix of students who had had the intervention in sixth grade, had been in control classrooms in sixth grade, and were new to the school.

Given this situation, we conduct two different comparisons for Year 2. For measures of knowledge of target words specific to seventh-grade instruction—namely, word knowledge, lexical access, and comprehension of text passages containing target words—we compare students who received RAVE in seventh grade with those in the control group in seventh grade. For measures tapping general skills—namely, morphological analysis and general comprehension—we compare students who had had two years of the intervention to students who had never received the intervention. Our rationale was that these different comparisons would best capture learning of the Grade 7 target words and distal effects that may have accumulated over the two years of the study.

Participants

One hundred and eight students were enrolled in the seventh grade. The school's seventh-grade structure placed 21 students in a separate honors class, while the other 87 students were heterogeneously grouped into four classrooms. As expected, there was a significant difference in GMRT at pretest between the five classrooms, $F(4, 103) = 14.53$, $p = .0001$, with post hoc comparisons showing differences between the honors section and each heterogeneously grouped classroom, with no significant differences between the other four classrooms. Thus, we selected as participants for Year 2 only students in the four heterogeneous classrooms ($n = 87$). Students in

two of those classrooms were designated to receive RAVE instruction (n = 44), and students in the other two were designated as control (n = 43). Of the 44 students in RAVE, 25 had received RAVE instruction in Year 1, 11 had been in the control condition, and 8 students were new to the school. Of the 43 control students, 19 had received RAVE in Year 1, 14 had been in the control condition, and 10 were new to the school.

One teacher, a European American female with 12 years of teaching experience, taught all classes. In seventh grade, five teachers each taught one subject to all students, providing students with daily 50-minute periods for reading, language arts, math, social studies, and science. RAVE was taught as part of the reading period, and lessons were, as in Year 1, 10 to 20 minutes long. Reading class focused on reading novels using study guides, which included some limited vocabulary work consisting of learning brief definitions for selected words from the novels. Control classes received a dosage of vocabulary instruction equal to RAVE students from a commercial vocabulary program (Bacon, 2005).

Materials

The seventh-grade RAVE intervention presented 96 AWL words in 16 seven-day cycles of instruction and four interspersed review cycles over about 22 weeks. The lessons followed the same general format as in sixth grade. The cycles were organized into themes, words were introduced with two contexts, definitions, and an activity to spur use of the words, followed by activities to promote active processing. For every other cycle, the final day presented a lesson in morphology, alternating with a review of the past two cycles. The morphology lessons again focused on Latin roots from words taught in that cycle. A total of 16 roots and 43 root-related words was taught.

Procedures

Procedures were the same as in Year 1. The research team provided the teacher with the same workshop as the sixth-grade teachers received and provided notebooks of all lessons. The teacher implemented the lessons daily. Members of the research team observed lessons and met with the teacher each cycle, keeping detailed field notes that were shared among the research team. A systematic observation protocol was used to collect fidelity data, as in Year 1, and all lessons were audio recorded and transcribed.

Measures

As in Year 1, we included four researcher-designed learning measures—a test of word knowledge for all taught words, a lexical decision task, a morphological awareness task, a text comprehension task—and a standardized comprehension task. Also as in Year 1, pretesting (EAV and GMRT) was conducted before instruction began in the fall. Posttesting, which included all

measures, was conducted in the two weeks following completion of instruction.

Word knowledge Test

A version of the EAV was developed for all 96 words taught in seventh grade. As described earlier, the EAV foils were created in an attempt to capture types of partial knowledge of words. We had predicted that semantic foils would be the most difficult and the unrelated and orthographic less difficult. Our prediction was correct for the semantic and unrelated foils, but the orthographic foils' performance was unstable (McKeown et al., 2014). Thus, we replaced orthographic foils with syntax foils, which were sentences in which all possible responses represented a different part of speech from the target word. For example, "Sarah wanted to _____ a better outfit for her job interview" for the target word, *passive*. We hypothesized these as easy for learners to reject, with even very shallow levels of word knowledge.

Lexical Decision Task

A lexical decision task was administered to assess fluency of access to target word meanings. Twenty RAVE words were selected from the seventh-grade program, which, as in Year 1, were the words with the highest scores at pretest for both RAVE and control students. Matched words and nonwords were selected following the Year 1 procedure.

Text Comprehension Measure

The passage comprehension task for seventh grade was based on an existing comprehension assessment, TOIW-C, used with fourth graders in an efficacy study of vocabulary instruction (Apthorp et al., 2012). It was administered to 1,450 fourth-grade students and demonstrated a significant difference between instructed and control students, with an effect size of 0.41, using Glass's *d*, and a reliability coefficient (Cronbach's α) of .75. Both versions of the task comprised four short passages, each containing six target words followed by six multiple-choice questions. Each question required an inference about the role of a target word in the passage. Neither the question stem nor answer choices contained the target word. For example, for our task, a text sentence stated: "They spent 20 years on the project, until finally they had a reliable model." A question geared to the word *reliable* read: "What were the Honda scientists able to achieve?" and the correct choice was, "They built a robot that can be counted on to perform well."

The passages for our measure were expository, wheras in Apthorp et al. (2012), the texts were narratives. The texts contained between 143 and 147 words each and had Flesch-Kincaid readability levels between 8.3 and 8.5.

Students read each passage independently and responded to the multiple-choice questions after each text.

Morphological Awareness Measure

A form of our dynamic assessment task was developed around roots taught in seventh grade, with each item containing a root-related word that had not been taught. The task was administered to 24 students (12 who experienced RAVE in both years, 12 who had not had RAVE in either year of the study). The 12 RAVE students were selected to represent low middle– to high middle–performing students on the GMRT; 12 comparison students were selected to match them. Students who were more than 1.5 standard deviations above or below the mean per national percentile rank on the GMRT were not included. There was no significant difference in reading achievement between students in the RAVE and control conditions in this subsample per an independent samples t test (t = 1.56, p = .590).

General Comprehension Measure

As in Year 1, the GMRT (Level 7/9, Forms S&T, MacGinitie et al., 2007) was administered pre and post to gauge potential influence of student learning on general comprehension. In particular, we examined results over two years for students who had been in RAVE both years (25 students) in comparison to students who had been in the control condition both years (14 students).

Administration of Measures

The EAV, GMRT, and text comprehension task were administered to all students in each classroom during their reading period. Because of problems with the school's computer lab machines, the lexical decision task was administered to all students in groups of six on laptops in an unoccupied classroom. The morphological awareness task was administered individually to the selected students.

Fidelity of Implementation

Fidelity of implementation was measured as in Year 1, with a transcript analysis of two full instructional cycles from all RAVE classes, once in the fall and once in the spring. Classroom observations to estimate instructional quality were conducted in both RAVE and control classes once in the fall.

Transcript analysis. Two members of the research team coded transcripts of all seven lessons in the two cycles for the percentage of scripted questions and explanations that were implemented.

Table 5

Means and Standard Deviations for Seventh-Grade Evaluation of Academic Vocabulary Word Knowledge Test Scores as a Function of Treatment

	n	Pretest	*SD*	Posttest	*SD*
Robust Academic Vocabulary Encounters (RAVE)	43	250.47	31.68	258.00	39.20
Control	42	241.36	35.49	230.86	39.98

Observations. In the fall, two members of the research team observed all participating classes using the same observation protocol used in Year 1. All observations were of the same RAVE lesson in each RAVE class and the same control lesson in the control classes. Reliability of coding was calculated as percent exact agreement. Overall exact agreement between observers was 94%.

Results

Analysis

As in Year 1, all analyses were conducted in SAS. The PROC MIXED or PROC GLM procedure, controlling for the Grade 7 GMRT pretest score. Comparisons were performed using the LSMEANS command with a Tukey adjustment. The assumption of sphericity was met in all cases and the assumption of normality was met except where noted.

Word Knowledge

A mixed analysis was performed on EAV test scores as a function of treatment and time controlling for GMRT pretest. There were two groups: RAVE (n = 43) and control (n = 42). The assumption of normality was met for all groups and timepoints except for the RAVE group at posttest. After controlling for GMRT pretest score, there was no significant interaction on EAV scores between treatment and time, $F(1, 36) = 2.14$, $p = .15$. While there was no significant main effect of time, $F(1, 36) = 2.57$, $p = .12$, there was a significant main effect of group, $F(1, 36) = 8.31$, $p = .007$, with the RAVE group having a higher score averaged across time compared to the control group (mean difference = 21.84) (see Table 5).

Lexical Decision

A mixed analysis was performed on average reaction time by word type controlling for GMRT pretest score. Word type (RAVE vs. matched) was the within-subjects variable, and group (RAVE vs. control) was the between-subjects variable. Only response times for correct decisions were included
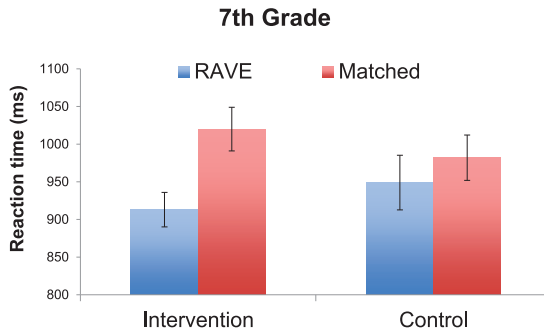
7th Grade

Figure 2. **Mean reaction time in milliseconds to Robust Academic Vocabulary Encounters (RAVE) and matched words for intervention and control students on lexical decision task in seventh grade.**

in the analysis. (Accuracy rate was 95.86%.) There was a main effect of word type such that RAVE words were responded to faster than matched words, $F(1, 82) = 17.14$, $p < .0001$, but there was no main effect of group, $F(1, 82) = .19$, $p = .67$. Importantly, there was a significant interaction between group and word type, $F(1, 82) = 5.02$, $p = .03$. Students in the RAVE group had significantly faster reaction times for RAVE compared to matched words, $t(82) = 4.68$, $p < .0001$, while the control group did not, $t(82) = 1.34$, $p = .57$ (see Figure 2).

Text Comprehension

A between-subjects GLM analysis was performed on passage comprehension scores as a function of treatment controlling for GMRT pretest. There were two groups: RAVE ($n = 44$) and control ($n = 43$). Analysis showed a difference in passage comprehension posttest score between groups that, while not significant at the .05 alpha level, was substantively meaningful as it indicated a trend in favor of the RAVE group, $F(1, 41.03) = 3.32$, $p = 0.07$, partial $\eta^2 = .03$ (see Table 6).

Morphological Awareness

A series of between-subjects GLM analyses were run to analyze performance on the recognition, meaning, and comprehension dimensions controlling for GMRT pretest. RAVE students scored significantly higher than the control group on the recognition, $F(1, 69.91)$, $p < .0001$, partial $\eta^2 = 0.65$; meaning, $F(1, 63.53) = 13.75$, $p = .001$, partial $\eta^2 = .040$; and comprehension dimensions, $F(1, 33.89) = 4.07$, $p = .05$, partial $\eta^2 = 0.19$,

Table 6
**Means and Standard Deviations for Seventh Grade Passage
Comprehension Score as a Function of Treatment**

	n	Mean	*SD*
Robust Academic Vocabulary Encounters (RAVE)	44	12.48	4.12
Control	43	10.93	3.63

Table 7
**Means and Standard Deviations for Each Dimension of the Morphological
Analysis Task Dynamic Assessment for Seventh Grade**

		RAVE			Control	
Dimension	*n*	*M*	*SD*	*n*	*M*	*SD*
Recognition	12	6.08	1.44	12	2.50	1.57
Meaning	12	4.75	2.30	12	1.50	1.88
Comprehension	11	11.55	3.17	9	8.22	3.35

Note. Recognition possible maximum = 9; meaning possible maximum = 18; comprehension possible maximum = 18. RAVE = Robust Academic Vocabulary Encounters.

demonstrating small (meaning and comprehension) to large (recognition) treatment effects in this small sample (see Table 7).

General Comprehension

To investigate the impact of RAVE on a general comprehension standardized test, a mixed analysis was performed on GMRT ESS as a function of time and treatment across the two years of the study. Two timepoints were chosen for this analysis because of the small number of students in the control or RAVE group for two consecutive years and the resultant lack of power to detect significant differences in a longitudinal analysis including all four timepoints. The within-subjects variable was time with two levels, pretest in sixth grade and posttest in seventh grade. The between-subjects variable was the pattern of intervention with two levels (assignment to the control group in both sixth and seventh grades or assignment to intervention in both sixth and seventh grades).

While there was no significant interaction between time and RAVE, $F(1, 66) = 0.42$, $p = .52$), there was a significant main effect of time, $F(1, 66) = 11.33$, $p = .001$, and while not significant at the .05 alpha level, a substantively notable trend in main effect between groups, $F(1, 66) = 3.27$, $p = 0.08$. For time, posttest scores were significantly higher than pretest scores (mean

Table 8
**Means and Standard Deviations at Two Timepoints
on Gates-MacGinitie Reading Test**

Group	*n*	Time	Mean	*SD*
Control–control	14	1	501.14	21.52
	14	2	509.14	31.67
RAVE–RAVE	25	1	506.67	19.25
	25	2	520.48	23.89

Note. Control–control = students assigned to control condition in sixth and seventh grades; RAVE–RAVE = students assigned to the RAVE condition in sixth and seventh grades; Time 1 = pretest in sixth grade; Time 2 = posttest in sixth grade; Time 3 = pretest in seventh grade; Time 4 = posttest in seventh grade. RAVE = Robust Academic Vocabulary Encounters.

difference = 10.90) averaged across groups, and scores for the RAVE group were significantly higher than for the control group averaged across time (mean difference = 8.44; Table 8).

Given the small sample size, the lack of sensitivity of large-scale standardized assessments to specific classroom-based interventions, the significant effect of time, and trending significance of group effect, LSMEANS comparisons with a Tukey adjustment were performed to explore differences that might exist. This examination revealed no significant differences in the pairwise comparisons except for significantly higher posttest scores compared to pretest scores for the RAVE students only, $t(66) = 3.47$, $p = .005$ (see Figure 3).

Fidelity of Implementation

For the transcript analysis, the scripted material was implemented 93% of the time across the two RAVE classes. For observations of instructional quality, scores revealed higher average levels of participation in RAVE classes and higher scores for timing and pacing of lessons in control classes. Results are presented in Table 4. Otherwise, scores for both RAVE and control classes reflected positive classroom climate and strong teacher preparedness to implement the lessons. Scores are presented as the average for each condition.

Discussion of Year 2

Results from implementation in seventh grade confirmed the sixth-grade results suggesting that RAVE is effective for promoting academic word knowledge and enhanced lexical access, both of which are associated with improved comprehension. Also as in sixth grade, RAVE students demonstrated greater ability to apply meanings of roots to infer meanings of unfamiliar words to comprehend sentences than control students, as measured by the morphological awareness task. In contrast to the sixth-grade
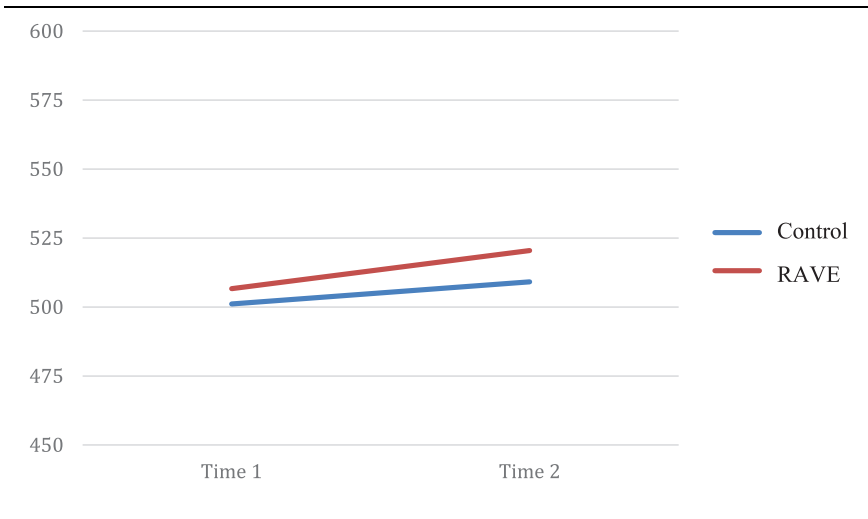
Figure 3. **Means at two timepoints on Gates-MacGinitie Reading Test. Time 1 is sixth-grade pretest; Time 2 is seventh-grade posttest. Control–control indicates participants who were in the control group for both sixth and seventh grades. RAVE–RAVE indicates students who were in the intervention group for both sixth and seventh grades. RAVE = Robust Academic Vocabulary Encounters.**

findings, results indicated a small advantage for RAVE on comprehension of text containing taught words. Effects on a standardized comprehension measure for the students who had experienced the intervention across two years were significant relative to students who had been in the control group both years. This result may suggest that general comprehension effects from vocabulary instruction take time to accumulate.

Examining Learning Processes Across Grades

In this section, we address Research Question 6, regarding what interactions during RAVE instruction demonstrate about the process of establishing word meaning. Our purpose with this analysis was to understand how students process elements of instruction toward building semantic representations. A key to the process of building an initial semantic representation is extracting relevant information from the resources available and integrating it into a coherent meaning for the word. Thus, we examined student interactions in lessons in which the words were introduced, and we focused on how students used the resources available—contexts and definition—to respond to the prompt that asked how the meaning of the newly introduced word fit the context.

Method of Analysis

We analyzed interactions for 25% of the contexts, which comprised all contexts from four cycles for each class, or 144 context discussions from the three sixth-grade classes and 96 context discussions from the two seventh-grade classes. Two cycles were selected from near the beginning of the year (Cycles 3 And 4) and two from near the end of the year (Cycles 9 and 10).

Each interaction was scored 0 to 2, with 0 indicating that integration had not occurred, and attention was focused solely on either the context or the definition; 1 indicating that effort was made to integrate the context and definition, but elements of the definition or context were used inappropriately or the response was too vague to trace to the relevant elements; and 2 indicating that context integration was achieved and elements of both the definition and the context were identifiable and used appropriately. Responses for each score are illustrated in Table 9. The scoring was developed collaboratively as the research team reviewed interactions from transcripts not included in the analysis. One team member then scored all interactions for the target contexts, and 20% were scored by a second team member for interrater reliability. Agreement of 86% was reached. Differences for the 20% were resolved through consensus.

Results

The results of our analysis indicated that context integration interactions were successful (score of 2) 74% of the time (65% in sixth and 85% in seventh), partially successful (score of 1) 16% of the time (19% in sixth and 14% in seventh), and unsuccessful or lacking (score of 0) 10% of the time (16% in sixth and 4% in seventh). Thus, the results show that students were able to integrate context and definition the majority of the time. However, given that students had explicitly designed context and definitional information available for completing the task, the results suggest that the process is not automatic.

Our review of students' interactions with the lesson contexts suggested three areas in which students appeared to face difficulty in their initial attempts to build a representation of a newly presented word: interpretation of elements of the definition, integrating elements of context and definition, and managing different senses of a word. Excerpts from lesson transcripts pertaining to each issue are examined in the following sections.

Interpretation of Definitional Elements

The definitions used in RAVE were designed around a core meaning to fit multiple senses of a word and illustrate typical constraints around the word's use. Even with such carefully constructed definitions, we found

Table 9
Scoring Categories and Examples for Integrating Meaning and Context

Score	Word	Context Summary	Definition	Integration Response
0	*Controversy*	There is controversy over an ingredient in toothpaste, Triclosan, which fights bacteria but may be unsafe and harm the environment.	A controversy is a strong disagreement about something that often involves a lot of discussion and angry feelings.	"'Cause they have to test it to see if will be harmful—which it is harmful if you have a cut in your mouth" (context only).
1	*Inclination*	A woman considers what to give to trick-or-treating children at Halloween; her inclination is to give out pretzels.	An inclination is a feeling that you are likely to act a certain way.	"She has a feeling that, um, they're going to act a certain way if she gives them, um, pretzels" (elements of definition not used appropriately).
2	*Distort*	News coverage distorts crime statistics.	If you distort something, you twist or alter it so that it looks or sounds different from the way it really is.	"That it doesn't . . . you don't understand what it really is and that it sort of changes the way it was, originally."

students exhibiting an issue highlighted by Scott and Nagy (1997), the tendency to interpret one word or phrase in the definition as the entire word meaning. For example, in a discussion of the word *facilitate*, defined as "If you *facilitate* an action or process, you do something to make it easier or more likely to happen," students were presented with a context about activities that could "facilitate learning." When asked how the meaning fit the context, a student responded, "Facilitate is like an action or process, so maybe they'll be thinking about it, so it would be, like, processing in their brain." Thus, the student interpreted the object of facilitation—"action or process"—as the word's meaning.

Another such example occurred with the word *criteria*, defined as "*criteria* are the list of requirements you use to make a decision or a judgment about something." After the class had read a context about criteria for being a blood donor, such as weighing at least 100 pounds, discussion proceeded as follows:

> Ms. H: Okay, so this context says that blood donors must meet certain criteria. What's that all about?
> Holly: Um, they want them to donate blood.
> Ms. H: Okay, that's true but what's the—what's the criteria?
> Holly: Uh, they need, yeah, they want them to make a decision.

Integrating Context and Definition

Bringing together relevant aspects of definition and context to describe how word meaning fit a context sometimes seemed a fragile point in establishing a word's meaning, as in the following examples. In the first example, the class is discussing a context for the word *generate*, which states that Tom's Shoe Company wants to "generate awareness" that many children need shoes. The definition is: "If you *generate* something, you cause it or produce it." Note that a student responds with a word from the definition but no connection to the context. After the teacher's follow-up question, the student responds with information from the context but loses the connection to the word *generate*.

> Ms. A: This context says Tom's Shoe Company wants to generate awareness that many children need shoes. What's that all about?
> Fer: Um, he wants to produce it.
> Ms. A: What do you mean?
> Fer: He wants, like, people to buy the shoes so other kids can have shoes.

Ms. A then called on a second student, who provides the integration:

> Ms. A: Um, yes, he wants to generate awareness. What does that mean?
> Jace: He wants to, like, produce awareness in people, like, he wants them to realize and, like, and around the world start realizing about this big thing.

The degree to which students struggled to integrate word meaning with the context varied considerably and often seemed to depend on the availability of a familiar word or phrase that could convey the role of the word in the sentence. For example, the word *sustain* was introduced and defined as "If you *sustain* something, you help to keep it going." The initial context presented for *sustain* described texting as a way to sustain friendships with people who have moved away. Students easily reframed that as "keep their friendships going."

Managing Different Senses

RAVE instruction purposefully focused on multiple senses. Managing different senses sometimes presented challenges to students, often for words that provided distinct concrete and abstract senses, such as the words *suspend* and *foundation*. Discussions of *foundation* in two classrooms, Ms. P's and Ms. H's, provide perspective on managing senses.

The word *foundation* was introduced with a context about damage from Hurricane Katrina, stating that the hurricane ripped houses off their foundations and that "people had to live in government trailers until new foundations were laid and new homes could be built on them." The following definition was then introduced: "A *foundation* provides support on which something else can be built." In both classes, discussion suggested that students were not clear about what a house's foundation was. The following excerpt is from Ms. P's room:

> Ms. P: Okay, so how does the meaning of foundation make sense in this context that we just read, Elana?
> Elana: Um, because, like, it says in the, um, definition . . . it means, like, to, like, build something else. And, it says, like, new homes . . . like, they had to live in, like, the government trailers until new homes were built in, like, the place where it happened.
> Ms. P: Mmm, hmm. So, who can tell me a little bit more about what that has to do with foundation? Maria?
> Maria: Um, the houses were destroyed and then so, like, they can help it build it again.
> Emma: All of it, um, well not all of them, but, some houses were destroyed by the wind and water and everything so, so people, would have to have, like, some people would donate and then that'd be like foundations for.

It appears that Emma's response, and possibly Maria's, refer to the sense of foundation as an organization that provides funding support, which we had not anticipated would be familiar to students. That sense of foundation may have been at play in Ms. H's class as well. When Ms. H posed the question of how the meaning of foundation made sense in the context, a student responded, "Because, the house has foundations that they're going to."

The second context for *foundation* introduced an abstract sense that portrayed "trust as the foundation of the relationship between a dog owner and his dog." The context involved a dog owner's belief that "with a foundation of trust, you and your dog can take on any challenge." Discussion in Ms. P's class showed evidence of understanding from one student:

> Ms. P: So, this context says that trust is the foundation of the dog-owner relationship. What's that all about, Dante?

> Dante: It means, like, um, trust is, like, what the relationship between the dog is, like . . . it's like a house that you build on the foundation. With (inaudible) the foundation is trust.

But there is evidence of remaining struggle, as well, for other students:

> Ms. P: Okay, and you build upon that, very good. Who heard what he said? Carly?
> Carly: He said, um, because then your dog can, it could learn anything if once you trust your dog and you have the um, I forget the word. (inaudible).
> Ms. P: The foundation?
> Carly: Yeah, at the foundation of, you can do anything with your dog.

Ms. P then provided an elaborated explanation uniting the ideas in the two contexts around the idea of foundation as "the support on which something else can be built." But when Ms. P asked students about their understanding across the two contexts, they seemed to have only the broad outlines of the word's meaning:

> Hannah: Like, like, you can build a trust, like, with a foundation.
> Ms. P: Okay, and?
> Hannah: Um, like, foundations of houses.
> Ms. P: Okay, no. you're, you're right. Can you tell me a little bit more about that?
> Hannah: Um . . .
> Ms. P: Can you add to that, Emma?
> Emma: To provide support.

These interactions were the initial encounters with a word, and students had many opportunities during the cycle to strengthen their concept of each word's meaning.

Discussion of Examining Learning Processes

The transcript excerpts presented in this section provide some perspective on the processes of learning academic words. Specifically, the three areas of learners' processing examined—interpretation of elements of the definition, integrating context and definition, and managing different senses of a word—demonstrate challenges students encounter as they attempt to build meaning representations across academic words and contexts. They also suggest that instructional interactions that prompt students to reflect on the information provided in definitions and context can support them in working through these word-learning challenges.

General Discussion

The RAVE intervention was designed around key aspects of word knowledge and word learning theory: polysemous senses of academic

words around a core meaning, exposure to patterns of use to enrich semantic representations and build fluent access, integration of words within contexts, and lexical morphology. The goal of instantiating these aspects into instructional activities was to support students in developing rich and flexible knowledge of academic words that could enhance comprehension of text. We used multiple measures to assess a progression of outcomes of students' learning. In this section, we discuss what the measures revealed about students' understanding and use of words. Finally, we consider limitations and implications of the study.

The most proximal measure, the EAV assessment of word knowledge, tapped students' ability to recognize correct uses and different senses of target words and reject incorrect uses. In both years of the study, RAVE students showed growth in knowledge of the target words relative to control students.

The lexical access and text comprehension tasks were considered intermediate measures of effects of instruction because they tapped the impact of vocabulary learning on a more general process—comprehension—yet they included instructed words. We view these kinds of tasks as potentially useful for advancing understanding of vocabulary instruction outcomes by suggesting when and how newly learned words begin to exert influence on text comprehension.

Lexical access results indicated that RAVE students developed more efficient semantic access to instructed words. This increased efficiency suggests that the words were more readily available for comprehension processing when encountered in text. The text comprehension tasks used in Year 1 and Year 2 were different. The Year 1 task required text recall, which is a global measure of comprehension, and the results showed that comprehension of text with taught words was not enhanced for RAVE students. The task used in Year 2 was redesigned to be a more direct measure of the influence of the vocabulary intervention because responding required integrating knowledge of instructed words into context in order to draw inferences about the text. The results of this task were at least suggestive of enhancement from the instruction.

The morphological awareness task and the GMRT were considered distal measures because they tapped general processes—morphological problem solving and comprehension—and did not include taught words. The results of the morphological awareness task suggested that instruction on lexical morphology had a positive impact on students' ability to use Latin roots to figure out meanings of novel words and use those words in comprehending sentences. This result adds to the growing body of literature suggesting that morphology instruction is worthwhile, which is significant because knowledge of specific roots can help students acquire new words independently and ability to analyze word parts can serve as a metalinguistic resource.

General comprehension gain is the most distal outcome but the ultimate goal of vocabulary instruction. Because we conducted our intervention over two years, we were able to examine longitudinal patterns of growth for a subgroup of students who remained in either the RAVE or control condition for both years. Although the number of students was small, results provided indication of an advantage for RAVE students in this exploratory analysis, suggesting that ongoing attention to vocabulary in a robust, interactive way may have a cumulative effect on general comprehension skill.

Analysis of lesson transcripts highlighted vocabulary-learning processes. The transcripts demonstrated the challenges of interpreting definitions, in particular where to focus attention within the definitional verbiage to grasp the essence of the word's meaning. The transcripts also suggested that integrating a newly learned word into a context is not a spontaneous process given students' difficulty selecting relevant information from definition or context and then combining it to make sense of the context. The examples of students' responses suggest how much information there is to manage in learning a new word.

Limitations and Implications

We acknowledge that limitations to the study led to only a partially successful intervention. Our small sample size and the setting within one school hinder the generalizability of our results. A related limitation was our inability to structure the RAVE and control groups over the two years of the study. Thus, although some students received the intervention in both sixth and seventh grades and some remained in the control group over both years, some students moved from RAVE to control or vice versa. And our group was limited the second year by the creation of the honors class. This situation limits our ability to draw implications about ongoing vocabulary instruction.

Our choice of assessments was deliberate but not without limitations. We chose to use primarily experimenter-designed measures because a major purpose of our study was to explore a range of effects of instruction and capture aspects of language processing that are not captured by extant measures. The advantage of such measures is that their relevance to the outcomes sought is greater than with standardized measures; however, the disadvantage is that the psychometric properties of such measures are not as well grounded.

We acknowledge that the results of our study are modest and may serve mainly to suggest directions for future investigations. Our study has implications for some key issues in vocabulary research through its approach to academic words, focus on morphology, and strategy for assessing outcomes. Academic words constitute vocabulary that students meet with some regularity in their academic texts, and their ubiquity could suggest that they might

be readily learned without explicit attention. Yet our results suggest that students do not develop strong knowledge of these words on their own.

Academic vocabulary is characterized by polysemous, abstract words that appear in a broad range of contexts. RAVE instruction addressed these qualities directly, explicitly introducing different senses of words and introducing words in pairs of contexts selected to represent different uses. Following word introduction, lessons used the words in various contexts that typified how students might encounter them in text. Our overall results suggest, and our transcript data demonstrate, that introducing academic words by providing typical contexts and directly confronting various senses and uses may be a useful approach to teaching these words. Although our transcripts indicate that students had difficulty with the prototypical characteristics of academic vocabulary, the transcripts also show that in the majority of cases, interactions around the contexts brought about fruitful results. Indeed, the difficulties that were revealed through addressing multiple senses and contexts suggest that if such word characteristics are not addressed, they may cause problems when students meet the words when reading on their own.

RAVE's focus on morphology also has implications for the field. Much research supports morphological awareness as a key metalinguistic ability. We demonstrated that brief lessons on lexical morphology, Latin roots, benefitted students' ability to comprehend sentences with novel words containing taught roots. Prior work has shown limited but growing support for instructional benefit and minimal focus on lexical, as opposed to derivational, morphology.

Finally, our study has implications for measuring the outcomes of vocabulary instruction. The range of assessments used in vocabulary research has been rather restricted (Elleman et al., 2009), and more detailed and specific understanding of how instruction contributes to students' growth is much needed. We assessed a range of outcomes across a continuum from proximal to distal effects. We placed our measures on the continuum based on the extent to which they directly assessed what was taught and the extent to which they tapped more general processes than those recruited in the instruction. How far along the continuum each measure lies is certainly arguable. For example, we acknowledge that our morphological measure was not as distal as a general comprehension measure. But we still consider it distal because students needed to engage in several levels of inferencing: Their task included inferring meanings of unfamiliar words based on a Latin root and inferring sentence meaning based on word meaning. Even though the roots within the words had been taught, the relationship between root meaning and word meaning is highly variable and thus not directly accessible from the root meaning.

We make no claim that our specific assessments were ideal, either individually or as a set. Indeed, the text comprehension measure used in Year 1 was problematic and was replaced for Year 2. Lexical access tasks are viewed as

indicating fluency of processes involved in comprehension, but they are not direct measures of comprehension. Yet because of the state of assessments of complex processes of word learning and comprehension, we see our efforts to assess a range of outcomes as a vital contribution to the vocabulary literature specifically and classroom intervention research more broadly.

Relatedly, we do not intend for our progression of measures to be exhaustive as to type of measure or process that should be evaluated. Rather, we view this work as a contribution to an ongoing conversation toward greater understanding of the role of vocabulary knowledge and vocabulary instruction. We believe that the exploration of a variety of measures is an important piece of that conversation as definitive results are rarely forthcoming from any individual study.

Notes

[1]The same covariate (Gates-MacGinitie Reading Test pretest) was used in all analyses in both Years 1 and 2 to ensure consistency of the modeling. While the covariate was a significant predictor in all Year 1 analyses except for the morphological awareness measure and all Year 2 analyses except for the lexical decision measure, there was no significant interaction of the covariate with any factor in any of the models for either year.

References

Adlof, S., Catts, H., & Little, T. (2006). Should the simple view of reading include a fluency component? *Reading and Writing*, *19*(9), 933–958.

American Institutes for Research. (2006). *Lessons and recommendations from the Alabama Reading Initiative: Sustaining focus on secondary reading*. Washington, DC: Author.

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society of Research in Child Development*, *58*(10), 1–186.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, *40*(3), 685–730.

Apthorp, H. S., Randel, B., Cherasaro, T., Clark, T., McKeown, M. G., & Beck, I. L. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness*, *5*(2), 160–188.

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. J. (2009). The impact of an instructional intervention on the science and language learning of middle grade English Language Learners. *Journal of Research on Educational Effectiveness*, *2*(4), 345–376.

Bacon, D. L. (2005). *Word master: Seeing and using words* (Level 7). Oakdale, NY: EDCON Publishing Group.

Baugh, A. C., & Cable, T. (1978). *A history of the English language* (6th ed.). New York, NY: Routledge.

Baumann, J. F., Kame'enui, E. J., & Ash, G. E. (2003). Research on vocabulary instruction: Voltaire redux. In J. Flood, D. Lapp, J. R. Squire, & J. M. Jensen (Eds.), *Handbook of research on teaching the English language arts* (pp. 752–785). Mahwah, NJ: Erlbaum & Associates.

Beck, I. L., Farr, R. C., Strickland, D. S., Ada, A. F., Hudson, R. F., McKeown, M. G., … Washington, J. A. (2009). *Storytown: A Harcourt reading/language arts program*. Orlando, FL: Harcourt, Inc.

Beck, I. L., & McKeown, M. G. (1985). Teaching vocabulary: Making the instruction fit the goal. *Educational Perspectives*, *23*(1), 11–15.

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

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

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). New York, NY: Guilford Press.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). The effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, *74*(4), 506–521.

Blanton, W., & Moorman, G. (1990). The presentation of reading lessons. *Reading Research and Instruction*, *29*(3), 35–55.

Bos, C. S., & Anders, P. L. (1990). Effects of interactive vocabulary instruction on the vocabulary learning and reading comprehension of junior-high learning disabled students. *Learning Disability Quarterly*, *13*(1), 31–42.

Bos, C. S., & Anders, P. L. (1992). Using interactive teaching and learning strategies to promote text comprehension and content learning for students with learning disabilities. *International Journal of Disability, Development and Education*, *39*(3), 225–238.

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, *80*(2), 144–179. doi:10.3102/0034654309359353

Burton, V. J., & Watkins, R. V. (2007). Measuring word learning: Dynamic versus static assessment of kindergarten vocabulary. *Journal of Communication Disorders*, *40*(5), 335–356.

Carlisle, J. F. (1995). Morphological awareness and early reading achievement. In L. Feldman (Ed.), *Morphological aspects of language processing* (pp. 189–209). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing: An Interdisciplinary Journal, 12*(3–4), 169–190.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*(2), 188–215.

Carroll, J. B. (1971). Development of native language skills beyond the early years. In C. Reed (Ed.), *The learning of language* (pp. 97–156). New York, NY: Appleton-Century-Crofts.

Connor, C. M., Spencer, M., Day, S. L., Giuliani, S., Ingebrand, S. W., McLean, L., & Morrison, F. J. (2014). Capturing the complexity: Content, type, and amount of instruction and quality of the classroom learning environment synergistically predict third graders' vocabulary and reading comprehension outcomes. *Journal of Educational Psychology*, *106*(3), 762–778.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Ruby, M., Crevecoeur, Y., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness*, *3*(2), 93–120.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*(2), 311–325.

Crosson, A. C., & McKeown, M. G. (2016). Middle school learners' use of Latin roots to infer the meaning of unfamiliar words. *Cognition and Instruction*, *34*(2), 1–24.

Curtis, M. E. (2006). The role of vocabulary instruction in adult basic education. In J. Comings, B. Garner, & C. Smith (Eds.), *Review of adult learning and literacy* (Vol. 6, pp. 43–69). Mahwah, NJ: Erlbaum.

Curtis, M. E., & Longo, A. M. (2001). Teaching vocabulary to adolescents to improve comprehension. *Reading Online*, *5*(4).

Davies, M. (2008). *The corpus of contemporary American English (COCA): 385 million words, 1990–present*. Retrieved from http://www.americancorpus.org

Davis, F. B. (1944). Fundamental factors in reading comprehension. *Psychometrika*, *9*(3), 185–197.

Durkin, D. (1978). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, *14*(4), 481–533.

E-Prime [Software]. (2002). Pittsburgh, PA: Psychology Software Tools Inc.

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

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied Linguistics*, *25*(1), 1–37.

Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, *17*(4), 257–285. doi:10.1080/10888438.2012.689791

Hairrell, A., Rupley, W., & Simmons, D. (2011). The state of vocabulary research. *Literacy Research and Instruction*, *50*(4), 253–271. doi:10.1080/19388071.2010.514036

Harmon, J. M., Hedrick, W. B., & Wood, K. D. (2005). Research on vocabulary instruction in the content areas: Implications for struggling readers. *Reading & Writing Quarterly*, *21*(3), 261–280.

Heller, R., & Greenleaf, C. L. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement*. Washington, DC: Alliance for Excellent Education.

Hiebert, E. H., & Cervetti, G. N. (2012). What differences in narrative and informational texts mean for the learning and instruction of vocabulary. In E. J. Kame'enui & J. F. Baumann (Eds.), *Vocabulary instruction: Research to practice* (2nd ed., pp. 322–344). New York, NY: Guilford Press.

Jenkins, J. R., Pany, D., & Schreck, J. (1978). *Vocabulary and reading comprehension: Instructional effects* (Tech. Rep. No. 100). Urbana, IL: University of Illinois, Center for the Study of Reading.

Kame'enui, E. J., Carnine, D. W., & Freschi, R. (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly*, *17*(3), 367–388.

Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. (2015). Word generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal*, *52*(4), 750–786.

Lesaux, N. K., Kieffer, M. J., Faller, E., & Kelley, J. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly*, *45*(2), 198–230.

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal*, *51*(6), 1159–1194. doi:10.3102/0002831214532165

Lubliner, S., & Hiebert, E. (2011). An analysis of English-Spanish cognates as a source of general academic language. *Bilingual Research Journal*, *34*(1), 76–93.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2007). *GMRT-MacGinitie Reading Tests, Level 7/9* (4th ed.). Rolling Meadows, IL: Riverside Publishing.

Margosein, C. M., Pascarella, E. T., & Pflaum, S. W. (1982, April). *The effects of instruction using semantic mapping on vocabulary and comprehension*. Paper presented at the meeting of the American Educational Research Association, New York, NY.

Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young at-risk children's word learning: A meta-analytic review. *Journal of Research on Educational Effectiveness*, *6*(3), 224–263.

Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2013). Literacy coaching to improve student reading achievement: A multi-level mediation model. *Learning and Instruction*, *25*, 35–48.

Matsumura, L. C., Slater, S., & Crosson, A. C. (2008). Classroom climate, rigorous instruction and curriculum, and students' interactions in urban middle schools. *Elementary School Journal*, *108*(4), 293–312.

McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking comprehension instruction: Comparing strategies and content instructional approaches. *Reading Research Quarterly*, *44*(3), 218–253.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior*, *15*(1), 3–18.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). *How specific aspects of vocabulary instruction contribute to the knowledge and use of words*. Paper presented at the American Educational Research Association Annual Meeting, Chicago, IL.

McKeown, M. G., Crosson, A. C., Beck, I. L., Artz, N. J., Sandora, C., & Moore, D. (2014, April). *Word knowledge and comprehension outcomes for the second year of implementation of an academic word vocabulary intervention*. Paper presented at the American Educational Research Association Annual Meeting, Philadelphia, PA.

McKeown, M. G., Deane, P. D., Scott, J. A., Krovetz, R., & Lawless, R. R. (2017). *Vocabulary assessment to support instruction: Building rich word-learning experiences*. New York, NY: Guilford Press.

Merriam-Webster.com. (2011). *Lexical*. Retrieved from https://www.merriam-webster.com/dictionary/lexical

Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, *53*(2), 253–279.

Myers, L. M. (1966). *The roots of modern English*. New York, NY: Little, Brown and Company.

Nagy, W. E., Berninger, V., Abbott, R., Vaughan, K., & Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second grade readers and at-risk fourth grade writers. *Journal of Educational Psychology*, *95*(4), 730–742.

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. B. Mosenthal, P. David Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 69–284). Mahwah, NJ: Erlbaum.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

National Center for Education Statistics. (2014). *The condition of education 2014 (NCES 2014-083), reading performance*. Retrieved from http://nces.ed.gov/programs/coe/indicator_cnb.asp

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific literature on reading and its implications for reading instruction* (NIH Pub. No. 00-4754). Washington, DC: National Institutes of Health.

Nelson, J. R., & Stage, S. A. (2007). Fostering the development of vocabulary knowledge and reading comprehension though contextually-based multiple meaning vocabulary instruction. *Education & Treatment of Children*, *30*(1), 1–22.

Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York, NY: Teachers College Press.

Pacheco, M. B., & Goodwin, A. P. (2013). Putting two and two together: Middle school students' morphological problem-solving strategies for unknown words. *Journal of Adolescent & Adult Literacy*, *56*(7), 541–553. doi:10.1002/jaal.181

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175.

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*(2), 282–296.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383.

Perfetti, C. A., & Adlof, S. M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how to assess reading ability* (pp. 3–20). Lanham, MD: Rowman & Littlefield Education.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–214). Philadelphia, PA: John Benjamins Publishing Company.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37.

Poehner, M. E., & Lantolf, J. P. (2010). Vygotsky's teaching-assessment dialectic and L2 education: The case for dynamic assessment. *Mind, Culture, and Activity*, *17*, 312–330. doi:10.1080/10749030903338509

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word-experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, *7*(1), 219–238.

Richter, T., Isberner, M., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading*, *17*(6), 415–434.

Scott, J., & Nagy, W. E. (1997). Understanding the definitions of unfamiliar verbs. *Reading Research Quarterly*, *32*(2), 184–200.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). *A practical guide to calculating Cohen's f², a measure of local effect size, from PROC MIXED*. Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2012.00111/full

Singer, H. (1965). A developmental model of speed of reading in Grade 3 through 6. *Reading Research Quarterly*, *1*(1), 29–49.

Snow, C., Lawrence, J., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, *2*(4), 325–344. doi:10.1080/19345740903167042

Spearritt, D. (1972). Identification of subskills of reading comprehension by maximum likelihood factor analysis. *Reading Research Quarterly*, *8*, 92–111.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, *56*(1), 72–110.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, *10*(4), 381–398.

Tippey, K. G., & Longnecker, M. T. (2016). *An ad hoc method for computing pseudo-effect size for mixed models*. Retrieved from http://www.scsug.org/wp-content/uploads/2016/11/Ad-Hoc-Method-for-Computing-Effect-Size-for-Mixed-Models_PROCEEDINGS-UPDATE-1.pdf

Vaughn, S., Martinez, L. R., Linan-Thompson, S., Reutebuch, C. K., Carlson, C. D., & Francis, D. J. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners. *Journal of Research on Educational Effectiveness*, *2*(4), 297–324.

Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly*, *52*(2), 203–226. doi:10.1002/rrq.163

Yang, C-L., Perfetti, C. A., & Schmalhofer, F. (2005). Less skilled comprehenders' ERPs show sluggish word-to-text integration processes. *Written Language & Literacy*, *8*(2), 233–257.

Yang, C-L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 55–89.