Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

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

Efficacy of Rich Vocabulary Instruction in Fourth- and Fifth-Grade Classrooms

Patricia F. Vadasy, Elizabeth A. Sanders & Becky Logan Herrera

To cite this article: Patricia F. Vadasy, Elizabeth A. Sanders & Becky Logan Herrera (2015) Efficacy of Rich Vocabulary Instruction in Fourth- and Fifth-Grade Classrooms, Journal of Research on Educational Effectiveness, 8:3, 325-365, DOI: 10.1080/19345747.2014.933495

To link to this article: https://doi.org/10.1080/19345747.2014.933495

Accepted author version posted online: 24 Jul 2014.
Published online: 02 Jul 2015.

Journal of Research on Educational Effectiveness, 8: 325–365, 2015

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345747.2014.933495

Routledge
Taylor & Francis Group

Efficacy of Rich Vocabulary Instruction in Fourth- and Fifth-Grade Classrooms

Patricia F. Vadasy
Oregon Research Institute, Seattle, Washington, USA

Elizabeth A. Sanders and Becky Logan Herrera
University of Washington, Seattle, Washington, USA

Abstract: A multi-cohort cluster randomized trial was conducted to estimate effects of rich vocabulary classroom instruction on vocabulary and reading comprehension. A total of 1,232 fourth- and fifth-grade students from 61 classrooms in 24 schools completed the study. Students received instruction in 140 Tier Two vocabulary words featured in two grade-level novels. Teachers were randomly assigned to either rich vocabulary (treatment) or to business as usual (control). Teachers in the treatment condition allotted 30 minutes per day to the intervention for 14 weeks. Hierarchical linear modeling revealed positive, significant treatment effects on distal and proximal measures of vocabulary and comprehension. However, average distal treatment effects were small (approximate d = .15) compared with proximal effects (approximate d = 1.24). Observations of teachers' language arts instruction indicated that treatment teachers spent significantly more time on vocabulary and less time on comprehension instruction than did teachers in the control condition. Results support the intensity and depth of the instruction for learning the taught corpus of words, and modest transfer to global vocabulary and comprehension.

Keywords: Vocabulary, efficacy, randomized trial, comprehension, proximal outcomes

INTRODUCTION

Vocabulary knowledge lies at an intersection of many reciprocal reading skills and experiences, culminating in a strong relationship between vocabulary and reading comprehension (Bolger, Balass, Landen, & Perfetti, 2008; Cunningham & Stanovich, 1997; Freebody & Anderson, 1983; Snow, Tabors, Nicholson, & Kurland, 1995). Multiple hypotheses have been proffered to explain this relationship (Anderson & Freebody, 1981; Freebody & Anderson, 1983; Mezynski, 1983; Perfetti & Stafura, 2014; Stanovich, 1986; Sternberg & Powell, 1983). Yet, as the National Reading Panel (2000) vocabulary subgroup noted, "Our knowledge of vocabulary acquisition exceeds our knowledge of pedagogy" (pp. 4–27). The differences in the vocabulary knowledge among students from different backgrounds and the low levels of reading achievement for many students in U.S. public schools underlie efforts to identify effective interventions to increase vocabulary knowledge (Biemiller, 2005; Biemiller & Boote, 2006; Kieffer, 2008; National Center for Education Statistics, 2012).

Extensive research on vocabulary acquisition suggests that students learn the majority of word meanings through natural reading contexts (Nagy & Anderson, 1984; Nagy &

Address correspondence to Patricia F. Vadasy, Oregon Research Institute, 150 Nickerson Street, Seattle, WA 98109, USA. E-mail: patriciav@ori.org

Herman, 1987). This research indicates that we should have realistic expectations for the direct or deliberate teaching of vocabulary (Nation, 2008). Further, as teaching demands increase it is not surprising that researchers have found that teachers dedicate limited time for vocabulary instruction (Blachowicz, 1987; Connor et al., 2014; Scott, Jamieson-Noel, & Asselin, 2003). Time during the school day allocated to direct vocabulary instruction must focus on teaching carefully chosen words and strategies (e.g., Nagy & Anderson, 1984). At the same time, research suggests that early vocabulary growth and background factors influence later vocabulary skills (Lee, 2011; Rowe, Raudenbush, & Goldin-Meadow, 2012) and that individual differences in vocabulary at school entry increase in size (Biemiller & Slonim, 2001; Cunningham & Stanovich, 1997). The role of word knowledge in reading comprehension and these individual differences in vocabulary knowledge have led researchers to develop interventions that successfully teach taught words as well as improve general vocabulary knowledge and reading comprehension.

Two foundational vocabulary intervention studies by Beck, Perfetti, and McKeown (1982) and McKeown, Beck, Omanson, and Perfetti (1983) have played a central role in research on vocabulary instruction. The studies describe an approach to teaching a specific group of vocabulary words used by mature language users that occur less frequently in daily language, are dispersed across domains, and allow for precision and specificity in language (see Nagy & Hiebert, 2011). This instruction seeks to develop depth of this word knowledge through interacting with words in varied contexts, expanding word relationships, affording opportunities for affective and cognitive responses to the words, multiple word exposures, and varied reasoning and practice activities. This instructional approach has come to be known as Rich Vocabulary (RVOC) instruction. The findings from these studies have informed recommendations about classroom vocabulary instruction (e.g., Kamil, 2004; Nagy & Scott, 2000; National Reading Panel, 2000; RAND Reading Study Group, 2002; Stahl & Nagy, 2006). In addition, aspects of the approach have been incorporated into multicomponent vocabulary interventions for both monolingual and English learner students (e.g., Loftus, Coyne, McCoach, Zipoli, & Pullen, 2010; Proctor et al., 2011; Puhalla, 2011). The approach has been outlined in two teacher handbooks (Beck, McKeown, & Kucan, 2002, 2008). In the three decades since the original RVOC studies were conducted, there have been many advances in the approaches that researchers take to evaluate reading interventions. More rigorous evidence standards account for the quality of research design including analytic approaches that account for nested data, moderators and mediators of effects, and that better characterize the magnitude of intervention effects for practitioners (see Lipsey et al., 2012). In this study we applied contemporary research design standards (Gersten et al., 2005; Institute of Education Sciences, 2003) to examine the efficacy of this widely used vocabulary practice. We attempted to replicate many aspects of the original research studies and implemented the study in classroom settings that reflect the teaching and learning demands that characterize today's public schools.

The original research on RVOC has influenced subsequent vocabulary research and instruction in two important ways. First, it prescribed a specific corpus of words for explicit instruction, and second it recommended instructional procedures to develop depth of vocabulary knowledge. Next we briefly review research on these two aspects of vocabulary instruction.

SELECTING WORDS FOR INSTRUCTION

Word frequency is one criterion used to select words that warrant direct instruction. High-frequency and root words are widely used in spoken English and in early reading texts.

Biemiller and his colleagues (Biemiller, 2005; Biemiller & Boote, 2006; Biemiller & Slonim, 2001) through a combination of direct testing and ratings of 6,000 root word meanings from Dale and O'Rourke's (1981) *Living Word Vocabulary* outlined a robust sequence in which *high-frequency root word* meanings are learned in Grades 1 to 6, identifying word meanings appropriate for explicit instruction. These high-frequency words are also often known as Tier One words that are learned easily by most students without direct instruction but may be less well known by at-risk and English learner students.

In contrast to Tier One words used in spoken language, another group of words are used primarily in written language and support learning in school contexts. In their original studies, Beck and colleagues targeted for instruction these words that are "judged likely to be unknown, yet useful and interesting" (McKeown et al., 1983, p. 6). These words were later identified as "Tier Two" words (Beck & McKeown, 1985): "words that are of high frequency for mature language learners and are found across a variety of domains" (Beck et al., 2002, p. 8); "words that are more sophisticated than the basic set but of high utility for literate language users" (Beck et al., 2008, p. 7). In contrast to Tier Three words that are labels for unfamiliar content and often associated with domain-specific knowledge (e.g., *photosynthesis, plebiscite, treble*), Tier Two words allow for making finer distinctions about commonly known concepts, like *sadness* (e.g., *grief, despair, depression, sorrow, bereft*).

ATTRIBUTES OF EFFECTIVE INSTRUCTION

The research on vocabulary instruction highlights several features associated with successful word learning. Instruction that is direct and that explicitly defines the vocabulary words is more effective than incidental instruction (Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009; Elley, 1989; Stahl, 1986). Other research supports instruction that combines this explicit with implicit methods of instruction (Marulis & Neuman, 2013; Stahl & Fairbanks, 1986). Multiple opportunities to encounter the words in the instruction, including hearing the words used in varied contexts, practice actively processing the words in speaking and writing contexts are also aspects of effective intensive instruction (Nagy & Scott, 2000; Stahl & Fairbanks, 1986; Stahl & Nagy, 2006). These features help develop the strong semantic representations that support reading comprehension (see Perfetti & Stafura, 2014).

The Vocabulary–Reading Comprehension Relationship

Reading comprehension is strongly influenced by lexical knowledge (Raymer, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). Verbal efficiency theory (Perfetti, 1992) suggests that rapid access to high-quality orthographic, phonological, and semantic word representations supports the reading fluency that allows individuals to allocate cognitive resources to comprehension. Verbal efficiency is limited when word meanings are not available for efficient lexical access and retrieval. Therefore, it has been proposed that vocabulary instruction that develops a rich and powerful network of meanings will affect general reading comprehension. Effective conditions include providing definitions and contextual information, multiple instructional encounters with the target words, and engaging students in active or deep processing (see Baumann, 2009; Graves, 1986). This type of instruction may prepare students to use their knowledge of academic language such as Tier Two words as tools for broader school learning (Nagy & Townsend, 2012), and perhaps greater vocabulary learning.

Numerous hypotheses have been proposed to explain the relationship between vocabulary and comprehension, and as Elleman, Lindo, Morphy, and Compton (2009) summarized, each is viable and they overlap. Reviews of vocabulary interventions that have examined this relationship have reported a similar pattern of relationships between instruction and comprehension: larger effect sizes (ESs) for experimenter-designed comprehension measures that feature taught words, and smaller ESs to support transfer to norm-referenced comprehension measures, and a similar pattern for vocabulary outcomes on proximal and distal measures (Elleman et al., 2009; National Reading Panel, 2000; Stahl & Fairbanks, 1986)

Finally, although vocabulary intervention research is growing, it remains difficult to identify specific features that achieve improvements in general vocabulary and reading comprehension. In the most detailed meta-analysis of vocabulary instruction for grades preK-12, Elleman et al. (2009) reviewed 37 studies of classroom vocabulary approaches that reported on passage reading comprehension outcomes. They found that vocabulary training had a strong overall positive effect on comprehension measured with custom (experimenter-designed) measures ($d = .50$) but only a small effect on norm-referenced comprehension measures ($d = .10$). They also found vocabulary instruction to be more beneficial in comprehension for students with reading problems ($d = 1.23$) than for average achieving students ($d = .39$). Elleman et al. (2009) were unable to detect differences in instructional characteristics accounting for variation in comprehension ESs. However, interventions that used higher levels of classroom discussion had larger impacts on vocabulary learning. Classroom discussions may provide opportunities to use and discuss vocabulary in meaningful written and spoken contexts (e.g., novel study), to receive feedback on correct word usage, and to hear the words used by others in varied contexts.

RICH VOCABULARY INSTRUCTION

In their 1982 study, Beck, Perfetti, and McKeown found that RVOC instruction had significant positive effects on students' reading comprehension and vocabulary knowledge measured on both experimenter-developed and norm-referenced tests of vocabulary accuracy (percentage correct) and latency (response time). Although the number of instructional encounters with target words varied in Beck et al. (1982) and McKeown et al. (1983), a minimum of 10 exposures was associated with significant treatment effects for comprehension. In Beck et al. (1982) there were no reliable differences in student knowledge of intensively exposed words and modestly exposed taught words, suggesting that their use of 10 to 18 exposures (median of 14) in the modest exposure condition was adequate to influence vocabulary and comprehension growth. Recently Baumann (2009) reviewed the Beck et al. (1982) study to estimate average Exposure Minutes of Instruction per word associated with vocabulary instruction that enhances comprehension and found that 14 average exposures per word produced a measurable positive influence on comprehension. Confidence in this dosage is somewhat limited because these early studies of rich vocabulary instruction (Beck et al., 1982; McKeown et al., 1983) featured small samples and quasi-experimental research designs that confounded classroom membership with students' treatment status. In contrast to contemporary classrooms, the early studies were conducted in classrooms with limited diversity in ethnic, language, and special education status, as well as initial reading ability. Analytic procedures did not routinely account for school, classroom, or student-level influences on outcomes.

Perhaps the most noteworthy finding of Beck et al.'s (1982) study was the significant effects of RVOC instruction on both proximal and distal vocabulary and comprehension outcomes. The effect on distal measures stands in contrast to meta-analyses that report large comprehension effects of vocabulary instruction are mainly confined to custom texts (Elleman et al., 2009; Pressley, Disney, & Anderson, 2007). The vocabulary subgroup of the National Reading Panel (2000) also found limited evidence that vocabulary instruction transferred to norm-referenced measures of reading comprehension. Several hypotheses about the vocabulary-comprehension link suggest that fluent access to word meanings is crucial for comprehension (Anderson & Freebody, 1981; Mezynski, 1983; Perfetti, 1985) and could explain why comprehension benefits were observed by Beck et al. (1982) on more proximal outcomes (i.e., custom texts that contained taught words), but it is harder to see how these hypotheses account for broader effects on standardized measures. Proximal and distal performance on comprehension may have benefited from the instructional tasks that are unique to Beck and associates' original interventions. These instructional tasks call for students to engage in deep reasoning about the target words in various contexts. Similar cognitive processes may undergird performance on both the custom and standardized measures, assuming that reasoning about word meanings and text meaning tap similar skills.

One recent study of an instantiation of the RVOC approach, called *Elements of Reading: Vocabulary* (Beck & McKeown, 2004), addressed limitations of the earlier research by implementing a cluster randomized trial with more than 16,000 students in kindergarten through fourth grade across 44 schools (Apthorp et al., 2012). This study sustained the supplemental 10- to 20-min daily intervention for 2 years to investigate cumulative impacts and included custom vocabulary and comprehension measures similar to those used in the present study. Significant treatment effects were found on the researcher-developed measures of vocabulary ($d = 1.24$) and comprehension ($d = 0.44$) at the end of Year 1 (4 months of intervention) for fourth-grade students most similar to the sample in the current study. However, treatment effects on standardized vocabulary and global comprehension measures (that do not feature taught words) were not found. Apthorp et al. (2012) echoed the questions raised in the recent reviews of vocabulary intervention studies, whether vocabulary knowledge is only one of many skills necessary for good comprehension which is also influenced by word reading, knowledge and use of comprehension strategies, and background knowledge. Finally, the researchers discuss their findings in the context of the wider debate concerning the limitations of standardized comprehension measures in detecting vocabulary growth (Elleman et al., 2009; Pearson, Hiebert, & Kamil, 2007).

There is broad consensus on the strong link between vocabulary knowledge and reading comprehension, and the effectiveness of features that characterize rich instruction in the knowledge of these useful and mature Tier Two words. This study builds on the early studies of the RVOC approach and examines the short-term effects of this established intensive vocabulary intervention on vocabulary and comprehension outcomes.

Research Questions

Our primary research question for this multicohort cluster randomized trial addresses the effects of RVOC instruction on fourth- and fifth-grade student outcomes relative to typical classroom (control) vocabulary instruction on (a) proximal and (b) distal measures of vocabulary and comprehension. Our second, auxiliary research question relates to how RVOC classroom teachers' allocations of content during their language/literacy blocks differed from typical time allocations in the control group.

METHOD

Participants and Random Assignment

Teachers

Each academic year for 3 years, fourth- and fifth-grade teachers in elementary schools in a major urban area of the Pacific Northwest were recruited in the late summer to participate in the study. Cohort 1 included 21 teachers (nine schools), Cohort 2 initially included 27 teachers (13 schools), and Cohort 3 included 14 teachers (eight schools). Six of the schools participated in more than one cohort (one in Cohort 1 and 2, four in Cohorts 2 and 3, and one in Cohorts 1 and 3), but no teacher or student participated in more than one cohort. In sum, there were 62 teachers from 24 schools that initially enrolled in the study. Each year of the study prior to student pretesting, teachers were randomly assigned (using a uniform random number generator in Microsoft Excel) to either treatment or control conditions: for Cohort 1, there were 11 treatment and 10 control teachers; for Cohort 2, there were initially 13 treatment and 14 control teachers (one treatment teacher dropped out of the study prior to treatment onset due to scheduling difficulties); and for Cohort 3, there were seven teachers in each condition.

Students

All students in each participating teacher's classroom were invited to participate each year. Information letters and blank consent forms written in English were given to teachers to distribute and collect, and translated versions were provided to teachers who requested them for any English learners in their classrooms. According to district records from the original combined set of 62 classrooms initially enrolled in the study, 1,540 students were available to participate ($n = 764$ from treatment classrooms and 776 from control classrooms), averaging 25 students per classroom for both groups (range = 12–35 students per class). Of these, consents were returned for 86% of control classroom students and 90% for treatment classroom students ($N = 1{,}354$ student consents returned, with $n = 689$ from treatment classrooms and 665 from control classrooms), with only 5% declining participation ($n = 63$; 35 treatment and 28 control students), and 1% returned too late to participate in testing ($n = 15$; five treatment and 10 control students). Of the 1,276 students with affirmative consents received by pretest, there was 3% attrition ($n = 44$ students, including 22 treatment and 22 control) due to one of the following: $n = 11$ absent for pretesting (seven treatment and four control), $n = 9$ removed from the study by the teacher due to those students missing the majority of the classroom language arts/literacy block (one treatment and eight controls, seven from one classroom with a large proportion of students receiving special education), $n = 8$ students from one treatment teacher in Cohort 2 who dropped out of the study before treatment onset, and $n = 16$ who moved before posttesting (nine treatment and seven control). In sum, 80% of students of participating classrooms completed the study (82% of treatment classroom students and 78% of control classroom students). A flowchart of general recruitment and attrition by cohort is given in Appendix A, Figure A1. *The final combined sample comprised* N = *1,232 students from 61 classrooms at 24 schools, with* n = *627 from 30 treatment classrooms and* n = *605 from 31 control classrooms.*

Table 1. Student and teacher sample characteristics

	RVOC		Control	
Characteristic	*N*	*(%)*	*N*	*(%)*
Students				
Female	317	(51%)	325	(54%)
Minority status	196	(31%)	220	(36%)
Asian	78	(12%)	94	(16%)
Black	61	(10%)	74	(12%)
Hispanic	26	(4%)	16	(3%)
Other	31	(5%)	36	(6%)
Grade level				
Fourth	272	(43%)	325	(54%)
Fifth	355	(57%)	280	(46%)
School services				
ELL	24	(4%)	33	(5%)
SPED	63	(10%)	50	(8%)
Teachers				
Female	23	(77%)	26	(84%)
Minority	4	(13%)	4	(13%)
Asian	2	(7%)	1	(3%)
Black	1	(3%)	2	(6%)
Hispanic	0	(0%)	1	(3%)
Other	1	(3%)	0	(0%)
Education				
Bachelor's	30	(100%)	31	(100%)
Master's	19	(63%)	22	(71%)
Certification				
State	27	(90%)	29	(94%)
National	3	(10%)	2	(6%)
Reading	2	(7%)	3	(10%)
ELL/ESL	2	(7%)	1	(3%)
SPED	1	(3%)	3	(10%)
Other	9	(30%)	9	(29%)

Note. $N = 1{,}232$ students (627 treatment, 605 control); $N = 61$ teachers (30 Rich Vocabulary [RVOC] treatment, 31 control). ELL = English language learners; ESL = English as a Second Language; SPED = special education.

Final Sample Demographic Characteristics

Table 1 displays demographic characteristics of the final combined sample of students and teachers. As can be seen, approximately one third of the students were minority status, half were from each grade level (fourth and fifth), and less than 10% were receiving English language learner (ELL) or special education services. The sample averaged $M = 10.25$ years old at pretest (range = 8.94–12.47 years, $SD = 0.60$), with no significant difference between conditions or cohorts (from a multilevel model of students within classrooms, within schools). Individual-level free and reduced lunch rates were not available for this study; however, school-level rates for 19 of the 24 sites were available and ranged from 8% to 69%

of students enrolled, with a median rate of 27% free and reduced lunch enrollment. Multilevel Bernoulli models (students within classrooms, within schools) showed no significant differences between groups or cohorts.

The teachers in the final sample were mostly female and nonminority, most had master's degrees, and most had at least a state certification in teaching. Cohort 1 had significantly more minority teachers (six of 21 teachers) compared with Cohort 3 (none of the 14 teachers), and Cohort 3 had significantly more teachers who had special education and reading certifications (three of 14) than Cohort 1 (none of the 21); no other differences were detected among teacher cohorts. Finally, there were no significant differences between treatment and control teachers (collapsed across cohorts) on any demographic characteristic, nor was there a significant difference between treatment and control conditions on years of teaching experience (treatment $M = 12.00$, $SD = 8.18$; control $M = 10.34$, $SD = 7.45$).

Treatment Instruction

Treatment Activities

The researchers developed RVOC treatment instructional materials using the guidelines provided in the most recent teacher handbook, *Creating Robust Vocabulary* by Beck et al. (2008). A set of activities were selected that fit into a 5-day weekly schedule, with 30 min of instruction on the first 4 days and a 10-min review quiz on the final day. A summary of the content of activities and reported teacher time spent on activities is provided in Table 2. To replicate the level of instructional intensity reported effective in the original research on the treatment instruction (i.e., a minimum 12 exposures per word), each word appeared in at least 13 instructional items (exposures) across the lesson activities each week of instruction. All necessary overheads, copies of student worksheets, and other materials were provided to teachers. All lesson components were scripted with the understanding that teachers would rephrase directions and prompts to better suit them and their students as needed. Finally, teachers completed daily attendance logs recording their completion of lesson activities.

Treatment Novels

The RVOC treatment instruction was developed to teach Tier Two vocabulary words we identified in two grade-level novels. The novel selection criteria were as follows: maximum length of 200 pages, reading and interest level matched for Grades 4 and 5, and a high number of Tier Two vocabulary words. The novels ultimately selected were *A Long Way from Chicago* (1998) by Richard Peck (a Newberry Honor Book) and *Maniac Magee* (1990) by Jerry Spinelli (a Newberry Medal Winner). Notably, *Maniac Magee* was one of the novels recommended for instruction in the Beck et al. (2008) handbook. Novel characteristics are provided in Table 3. Each novel was divided into seven weekly reading assignments that ranged from 15 to 30 pages. Teachers were allowed to assign the reading in the format of their choice to fit with the reading level and load of their students (e.g., read-aloud, partner reading, independent reading in class or as homework). Along with each book, students were provided a bookmark that listed the words and their corresponding page numbers.

Table 2. RVOC treatment instruction weekly activities by day

Day	Activity	Description	Example	Teacher Log Data		
				M	*(SD)*	*Adj M*
1	Definitions[a] (10 words/week)	Provide definition, students fill in missing words on worksheet	A *blunder* is a silly or careless mistake.	—		
	Examples[a]	Provide sentence that demonstrates word meaning then ask for a student example	Leaving your car unlocked and having your camera stolen is a *blunder*. Can you think of a *blunder* you have made?	—		
	Idea Substitution*,[a]	Provide a sentence that students rephrase to include target word	The first baseman's silly mistake cost the team the game.	—		
	Word Relations*,[b]	Ask whether/how the target word fits with another word	How do infamous and blunder fit together?	—		
	Across Day 1 Activities (repeated for two sets of five words/set, 15 min/set)			32.43	(2.94)	32.43
2	Synonym Matching[b]	Teams match two synonyms on cards to target words	*blunder:* (a) mistake, (b) blooper	10.91	(2.57)	10.90
	Sentence Stems[b]	Provide sentence with target word, students fill in the end	Her *blunder* made her feel _____.	10.46	(2.23)	10.45
	Example Selection[b]	Provide prompt with target word, students select which alternative matches	A mistake is more likely to be called a blunder if you make it when you're: (a) focused or (b) angry? Why?	11.26	(1.79)	11.23
	Across Day 2 Activities			32.62	(4.64)	32.49
3	Novel Context Question[c]	Review surrounding plot, read quote, then ask multiple choice question	Why was it assumed that ending up on Finsterwald's property was a *blunder*? Explain your answer choice.	17.62	(3.10)	17.60
	Occupation/ Role Prompt*,[c]	Different groups of students answer the same questions from a different perspective	How might a cook/musician/ teacher/ basketball player: Make a *blunder*? Become infamous? Test a theory?	14.46	(2.40)	14.26

(Continued on next page)

Table 2. RVOC treatment instruction weekly activities by day *(Continued)*

				Teacher Log Data		
Day	Activity	Description	Example	M	(SD)	Adj M
	Example Descriptions*,c	Describe examples of the target word	When are you most likely to make a *blunder*?	—		
		Across Day 3 Activities		32.08	(4.14)	31.84
4	Definition Match[a]	Students match words and definitions	*blunder* - careless mistake	8.34	(2.16)	8.30
	Sentence Writing*,c	Students write 2 to 3 sentences in response to a general prompt or a prompt related to the novel	Describe a *blunder* you made, and how it made you feel, and why it was a *blunder*.	11.90	(2.80)	11.89
	Paragraph Writing*,c	Students write a paragraph using several of the vocabulary words in their log	Write paragraph in response to novel prompt using vocabulary words.	14.24	(3.84)	14.24
		Across Day 4 Activities		34.48	(5.99)	34.47
5	Quiz	Complete and review multiple-choice quiz	A *blunder* is a (a) mistake, (b) apology, (c) thought, (d) loud noise	13.15	(6.34)	13.14
	Chapter Preview	Assign pages for next section, present background knowledge				—
		Weekly Across Days (Except Day 5 Chapter Preview)		144.75	(16.56)	144.12

Note. About half the words were featured in activities with an asterisk, and the remaining words were featured in the activities without an asterisk so that all words were used in the same number of activities. For activities that teachers recorded in log files, we report observed (unadjusted) means and standard deviations of activities as well as means adjusted for classroom membership from two-level random intercept models ($N = 30$ Rich Vocabulary [RVOC] treatment teachers across 20 schools; median intraclass correlation = .18). [a]Word meanings. [b]Word relations. [c]Contextualizing.

Table 3. Rich vocabulary treatment novel characteristics

Novel Title	Pages	Grade Interest	Grade Level	Lexile	DRA	Guided Reading
A Long Way from Chicago	148	6–8	4.6	750L	50	V
Maniac Magee	184	4–8	5.4	820L	60	W

Note. Grade-level information from *Scholastic Book Wizard*. DRA = Developmental Reading Assessment.

Treatment Words

Researchers identified 280 potential Tier Two words in the two novels and tabulated their frequency data (for uninflected word forms) from Zeno, Ivens, Millard, and Duvvuri (1995). For each weekly RVOC reading assignment, for each novel, we identified 10 words from this pool of words that were relevant to novel themes, of high interest to fourth- and fifth-grade students, and amenable to the treatment instructional format. The final word corpus included 140 words (see Appendix A, Table A1).

Treatment Training

Research staff provided training in the RVOC treatment instruction materials and procedures in two meetings with teachers before the intervention began, and in a follow-up meeting shortly after teachers began instruction. At the first meeting the researchers delivered materials and met briefly to orient teachers to the instruction. Teachers were instructed to review the manual, lessons, and student materials before the next training meeting. Research staff next scheduled individual or small-group training sessions that were about 1.5 hr long in which they previewed the teacher and student materials and reviewed instructional delivery techniques (eliciting student responses, giving feedback, and pacing) as well as the full 5-day cycle of lessons (each daily activity was explained and modeled). In the first 2 weeks of treatment instruction, a follow-up meeting was conducted either in person or via phone call. If requested, research staff conducted informal observations for teachers who needed more detailed feedback.

Treatment Fidelity

We adopted a multipronged approach to ensure extremely high levels of RVOC treatment fidelity, including initial site visits for procedural coaching and modeling, daily self-report logs of time spent on specific RVOC activities, teacher–researcher e-mail communications, and formal classroom observations of time spent on language arts/literacy content.

Site Visits and Self-Report Logs. For each cohort, during the first 2 weeks that treatment teachers were implementing RVOC lessons, researchers visited each classroom to ensure that each teacher was implementing lesson procedures (and to provide coaching and modeling as needed). Further, researchers used ongoing electronic and phone communications with teachers throughout the treatment to address questions or technical issues about lessons. Teachers were also trained to record their start and stop times each day of the RVOC lessons on an activity log. These daily logs were generally organized by lesson activity (exceptions were Day 1 and Day 5; Day 1 activities were integrated tasks, and Day

5 logs only the quiz task). Data from these logs were entered into a database that calculated the amount of time spent on the daily lesson activities, and descriptive statistics averaged across the treatment period are displayed in Table 2 (we also report adjusted estimated means, after controlling for teachers' school membership). As can be seen across all days, treatment teachers generally spent more than 30 min on lesson activities Days 1 to 4 and more than 10 min on the Day 5 quiz activity (approximately 8% more than the prescribed 30 min for Day 1 activities, 9% more on Day 2, 7% more on Day 3, 15% more on Day 4, and 32% more on the Day 5 quiz). More specifically, only two teachers reported spending less than an average of 30 min on Day 1 activities across the 14 weeks, and eight teachers reported spending less than an average of 30 min on Day 2 and 3 activities and less than the prescribed 10 min on the Day 5 quiz activity.

Classroom Observations. Treatment fidelity was also measured with three onsite observations of each teacher's entire language arts/literacy block, beginning the first week of treatment and ending the last week of treatment (see next section for detailed observation procedures). Because teachers implemented the lessons during their regular classroom language arts/literacy blocks, 30 min of the total block time should have been spent on treatment lesson activities. Lesson activities, described in Table 2, were coded by the third author to match three of four widely recommended features of vocabulary instruction: word meanings, word relationships, and contextualizing words (see Stahl & Nagy, 2006). Of the total instruction time allocated for Day 1 to 4 activities, lesson design was planned so that word meanings (developing depth of meaning) would take, on average, about 38% of treatment time (11 min), word relationships about 8% (3 min), and word contexts (contextualizing) about 54% (16 min). Further, time using treatment texts and lessons was also measured to gauge teacher pacing through the lessons as well as possible control group contamination (because classroom random assignment was across schools rather than within schools).

Classroom Literacy Instruction Observations (Treatment and Control Conditions)

Classroom language arts/literacy block observations conducted on teachers in each cohort served three important purposes: (a) to measure treatment teachers' fidelity to assigned vocabulary instruction (see earlier), (b) to measure any treatment contamination in control teachers' instruction, and (c) to measure control teachers' business-as-usual regular language arts/literacy block instruction. All observations were conducted for the teacher's entire language arts/literacy block to ensure comparability of content time estimates, and as such, it should be noted that treatment teachers' estimates included *but were not limited to* lesson time. Three observations provided stable instructional estimates and allowed us to test for any linear changes in teachers' instruction during the study.

Procedures

The instrument used to collect classroom teacher observation data is an adapted version of the *Instructional Content Emphasis–Revised* (Edmonds & Briggs, 2003). Trained observers visited each participating fourth- and fifth-grade classroom three times between fall and spring, at nearly equidistant occasions (after student pretesting and treatment onset) in the fall, in midwinter, and before posttesting in the spring. Using electronic timing devices, observers began recording time (hours: minutes: seconds) along with the appropriate

numerical codes corresponding to instructional characteristics, beginning at the onset of the language arts/literacy block. Thereafter, every time the teacher switched instructional characteristics, the time and corresponding numerical codes were recorded. To avoid potential drift, observers used timers so that they were reminded every 5 min to check whether the instructional code had changed. Recordings were made on a log sheet that included the classroom name, observer name, date, time of day, and multiple pages of rows for writing time and code entries. Data were then entered by a trained research assistant into an Excel database that corresponded to the log sheet entries. Time spent on each instructional code, for each teacher and observation occasion, was computed automatically with a subtraction function in Excel for each row in the database, and Excel pivot tables computed the sum of the time spent (all time calculations were scaled in minutes) across entries for each instructional code, by teacher and observation period. As such, time spent in specific literacy content areas for the entire duration of each teacher's language arts/literacy block was quantified.

Content Codes

Specific Instructional Content Emphasis–Revised content codes for this study included six mutually exclusive instructional areas: *Word Study* (e.g., word reading, spelling, and letter/sound relationship study), *Text Reading* (e.g., independent silent reading, and teacher read aloud), *Writing* (e.g., composition, grammar, and punctuation), *Vocabulary* (word meanings, word relationships, and processing words in contexts (contextualizing) and morphological analysis (which was not a specific component of the treatment lessons), *Comprehension* (e.g., comprehension strategies, monitoring of reading/listening comprehension), and *Other* (e.g., nonliteracy instruction such as behavior issues, transitions, or math content). (A copy of the fully described codes is available from the authors upon request.) Finally, we also measured the time spent on use of treatment lessons and texts, which overlapped with the time on the content codes. Quality of instruction was not recorded.

Observer Reliability

Three observers (two certificated teachers and the third author) conducted all classroom observations for all cohorts. Prior to the onset of each year of intervention, observers studied the coding instructions and participated in two formal training sessions. To establish reliability prior to onsite data collection, for each cohort, each observer independently coded 10 videotapes of classroom literacy instruction (each with different teachers/students) from Grades 2 to 5 that featured the primary literacy content codes. Data from these videotape codings were entered into an Excel database where time on content was automatically calculated by videotape, by observer (described earlier in more detail in the observation procedures). These data were then imported into SPSS in which internal consistency was estimated across observers (Cronbach's alpha), for each content code separately (i.e., treating videotapes as subjects and observers as variables). For each cohort, prior to onsite observations, observers were required to correlate at .70 or higher on each code. Across cohorts, the median reliability for each content code was .99, with the exception of comprehension (.98), other (.92), and treatment text use (.71). Across codes, the median reliability for Cohort 1 was .98, for Cohort 2 was .99, and for Cohort 3, .99.

Estimation of Content Means

As already described, we tabulated time (in minutes) spent on each content code for each teacher and observation in Excel. Table 4 displays the observed (unadjusted) means and standard deviations of the six content codes, four vocabulary content subcodes, and treatment-related content for each observation occasion as well as grand means across measurement occasions. Data were imported into HLM and three-level linear growth models were estimated (measurement occasions within teachers, within schools). Those results showed that teachers (a) generally spent significant (nonzero) amounts of time in the fall (intercept) on each content area, and (b) time allocations for each content area as well as vocabulary content subcodes remained constant during the intervention period, with the exception of writing instruction, which was estimated as increasing by 1.14 daily minutes/month from the initial fall observation of 2.70 daily minutes. We caution overinterpretation of this finding given that most classrooms also had separate writing blocks not observed for this study. Last but not least, as described in our forthcoming model results for our second research question (see Results section), comparison of the treatment and control teachers' intercepts or slopes for time spent on treatment lessons and texts showed no significant differences between groups (no control teachers showed use of the RVOC treatment lessons, although one control teacher had been using one of the treatment novels ad hoc in her instruction), indicating no evidence of control group contamination. More detail regarding the observation growth models can be obtained from the authors upon request.

Student Assessments

Each year, students were pretested in fall and posttested in spring, approximately 6 months apart. Students were tested in whole-class format by trained testers unaware of experimental group assignment. To achieve comparability with previous research, raw scores were used for all analyses. The following measures were administered using identical test forms at pretest and posttest.

Vocabulary was assessed using two experimenter-developed curriculum-based (proximal) measures (CBMs) and a norm-referenced (distal) measure as follows.

1. *Experimenter-developed CBM-Words (Proximal)*. To measure proximal vocabulary and comprehension, we constructed three CBMs that closely mirrored the assessments described in Beck et al. (1982) and McKeown et al. (1983), including taught word vocabulary (CBM-Words), sentence verification (CBM-Sentences), and story recall (CBM-Comprehension). The first two were vocabulary tasks, whereas the last was a comprehension task. Each of these measures was based on the corpus of 140 words selected from the two novels and included in instruction, although the use of the words varied depending on the measure. To select specific words for each measure, words were stratified by the 14 weeks of treatment instruction in which they were used, and within each week stratified by the relative frequency (Standard Frequency Index [SFI]) of the word (median split into low and high frequency). For the first two measures (taught word vocabulary and sentence verification), a single word was randomly drawn from within each of the 28 strata (56 words). The remaining words not sampled (140 total – 56 = 84) were used to create the third measure (story recall; see Comprehension next). This process helped ensure that each word was used only in one measure and that

Table 4. Observed classroom language arts/literacy block content minutes

Language Arts/ Literacy Block Content	RVOC Treatment Teachers[a]								Control Teachers[b]							
	Fall		Winter		Spring		*Mean*		Fall		Winter		Spring		*Mean*	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
General																
Word study instr	3.56	(8.43)	3.13	(8.23)	3.06	(6.62)	3.25	(6.03)	4.05	(7.69)	4.09	(6.64)	3.14	(6.93)	3.76	(5.94)
Text reading instr	24.14	(12.93)	24.12	(16.94)	24.31	(17.97)	24.19	(13.58)	27.00	(13.72)	24.29	(15.64)	22.11	(15.36)	24.47	(11.18)
Writing instr	1.95	(4.94)	4.72	(11.42)	6.41	(13.82)	4.36	(9.22)	2.88	(7.18)	5.19	(10.99)	4.39	(8.80)	4.15	(6.63)
Vocab instr	35.49	(12.44)	36.30	(13.13)	32.17	(10.91)	34.65	(8.82)	8.83	(9.10)	10.25	(11.89)	8.86	(10.26)	9.31	(7.16)
Word meanings[c]	12.11	(10.29)	11.97	(10.33)	14.97	(11.55)	13.02	(5.40)	4.99	(6.79)	4.52	(6.15)	4.96	(6.42)	4.82	(4.17)
Word relations[c]	1.59	(5.07)	1.71	(4.60)	1.32	(2.21)	1.54	(2.13)	1.29	(4.33)	0.56	(1.82)	0.09	(0.35)	0.65	(1.53)
Contextualizing[c]	20.27	(12.12)	21.84	(16.74)	15.46	(12.62)	19.19	(8.73)	2.34	(4.16)	4.39	(9.67)	3.22	(7.34)	3.32	(4.33)
Morph analysis	1.27	(5.79)	0.71	(3.52)	0.25	(0.64)	0.74	(3.07)	0.21	(0.75)	0.78	(1.84)	0.48	(1.75)	0.49	(0.85)
Comprehension instr	16.39	(15.61)	12.34	(10.38)	10.79	(10.86)	13.18	(8.20)	27.29	(16.84)	22.90	(15.27)	24.69	(17.33)	24.96	(12.49)
Other non-lit instr	13.25	(8.56)	13.12	(10.76)	11.54	(7.51)	12.64	(7.08)	10.69	(6.84)	12.97	(8.03)	10.97	(5.54)	11.54	(5.43)
Treatment specific																
Lesson use	33.27	(11.58)	31.26	(9.83)	30.85	(11.97)	31.79	(6.59)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Text use	6.86	(10.12)	8.57	(12.16)	7.84	(10.59)	7.76	(7.57)	0.00	(0.00)	0.69	(3.56)	0.20	(1.11)	0.33	(1.57)

Note. All values in daily minutes of language arts/literacy instruction (instr) blocks. Time between fall and winter observations = 1.5 months, and time between winter and spring = 1.2 months.

[a]$n = 30$. [b]$n = 31$. [c]General content contained in Rich Vocabulary (RVOC) treatment instruction. Multilevel linear growth models showed no significant linear changes over time in content times except for writing instruction (increase) but did show that RVOC treatment teachers had significantly greater time spent on vocabulary instruction and treatment-specific lesson and text use, but that control teachers had significantly greater time spent on comprehension instruction.

Table 5. RVOC treatment and CBM measure word characteristics

	Class				SFI		U	
Characteristic	*N*	*Noun*	*Verb*	*Adj*	*M*	*(SD)*	*M*	*(SD)*
RVOC Treatment Instruction	140	34	53	53	40.3	(8.10)	4.1	(9.60)
CBM Pretest/Posttest Measures								
Words	28	10	11	7	39.3	(7.40)	2.1	(3.00)
Sentences	28	6	8	14	40.8	(8.40)	5.2	(10.60)
Comprehension	28	6	11	11	36.6	(9.00)	4.2	(16.20)

Note. SFI and U from Zeno et al. (1995). RVOC = Rich Vocabulary; CBM = curriculum-based measure; SFI = Standard Frequency Index; logarithmic transformation of occurrences per million; U = occurrences per million; Adj = adjective.

each of the three measures had similar word characteristics (i.e., were representative of all weeks of instruction and overall SFI levels). Table 5 displays the word characteristics for each experimenter-designed measure as well as for treatment instruction words. The first task, a 28-item word vocabulary measure, included a vocabulary word stem, along with four possible response definitions for the word. The words were presented in minimal-context incomplete sentences that require a synonym to be selected to finish the sentence, for example, A blunder is a: (a) painful injury, (b) careless mistake, (c) historic event, (d) difficult task. To minimize the effects of decoding skill on the assessment scores, items were read aloud to the class while students followed along and individually selected a response on their copy of the task. Sample internal consistencies for all experimenter-developed measures are reported in the assessment descriptives table (see Results section), and each of the CBM forms are given in Appendix B.

2. *Experimenter-developed CBM-Sentences (Proximal)*. For the 28-item sentence-verification measure (CBM-Sentences), the tester asked a question about a vocabulary word that required the student to select a yes/no response. Fourteen of the items had a "yes" as the correct response (e.g., If something is *vast*, is it large?) and 14 had a "no" as the correct response (e.g., If you are *fascinated*, are you bored?). Again, to minimize the effects of decoding skill on the assessment scores, items were read aloud to the class while students followed along and individually selected a response on their copy of the task. This assessment is similar in design to the orally group-administered Test of Instructed Word Knowledge in Vocabulary used in the Apthorp et al. (2012) study.
3. *Norm-referenced measure (Distal)*. General vocabulary was measured using the Iowa Test of Basic Skills (ITBS; Hoover, Dunbar, & Frisbie, 2003) vocabulary subtest, which requires students to read a word silently in context and then select from among four choices the correct definition of the word. Target and distracter words were drawn from The Living Word Vocabulary (Dale & O'Rourke, 1981). Nouns, verbs, and modifiers are given approximately equal representation. Because we found no significant differences in word characteristics between the fourth- and fifth-grade levels for the ITBS Form A, Level 11, we used the fifth-grade test form for all students, which includes 27 items. According to the test manual, KR-20 reliabilities (internal consistencies) average .90 and .88 for fourth and fifth graders, respectively.

Comprehension was assessed using one experimenter-developed (proximal) CBM and one norm-referenced (distal) measure as follows.

1. *Experimenter-developed CBM-Comprehension (Proximal)*. Proximal comprehension growth was measured with the experimenter-developed story recall task (CBM-Comprehension). This task required students to read a story written with taught words embedded. The story was designed to contain a ratio of one treatment instruction word per 11 noninstruction words (words were selected from the pool of 84 words randomly assigned for this task). The story (*Horses for Rent*) was 498 words in length and contained 28 unique treatment instruction words that were used one or two times each for a total of 44 occurrences (Flesch-Kincaid grade level of 6.4). The students were given 13 min to independently (silently) read and answer 12 multiple-choice comprehension questions. Sample reliabilities are provided in the Results section.
2. *Norm-referenced measure (Distal).* The ITBS reading comprehension was used to measure general comprehension ability. This subtest required that students silently read 8 passages that varied in content and genre (e.g., biography, poetry, and animal behavior) and respond to questions about each passage's main ideas, literal details, and potential inferences that can be drawn using the passage information. Again, for simplicity, we used the 43-item fifth-grade form (Form A, Level 11) with all participating students. According to the test manuals, KR-20 reliabilities (internal consistencies) average .89 and .90 for fourth and fifth graders, respectively.

Analytic Approach

Due to the nesting structures (nonindependence) present in the research design, a multilevel hierarchical modeling approach was adopted for testing each of the research questions.

Student Outcomes

To test RVOC treatment effects on student outcomes directly relating to our research questions, we again employed three-level hierarchical models (students nested within teachers, within schools) for both pretests and posttests using full information maximum likelihood in HLM7 software. We began with preliminary intercept-only (i.e., "empty" or "unconditional") models to estimate the dependence at each level of the data to confirm that each level should be taken into account, and indeed we found a pattern of nonindependence for both schools as well as classrooms on student outcomes. As another preliminary step, we tested cohort effects using two effect-coded classroom Level 2 variables for the three cohorts and found that there were no such effects on any measure, at pretest or posttest.

Next, we estimated RVOC treatment effects on pretests to determine whether pretest should be controlled for in our subsequent posttest analyses. These analyses showed that the RVOC treatment group was consistently lower than controls at pretest despite random assignment, both at the student level and the aggregate classroom level. The approximate difference between groups on pretests at the student level, using three-level model estimates, averaged 0.22 standard deviations favoring the control group (range = 0.19–0.25), and the difference between groups at the aggregate classroom level, using two-level model estimates, averaged 0.46 standard deviations favoring the control group (range = 0.37–0.56). As such, respective pretests were used in all posttest analyses to control for these baseline differences between groups as well as maximize statistical power. More specifically, we used both individual student-level pretest (standardized *within* classroom) and classroom-level aggregate pretest (standardized *across* classrooms) in our posttest models.

Finally, we estimated whether there were any grade-level effects on posttests after accounting for pretest using an effect-coded variable at the student level and found no grade level effects. Note that ELL status was not incorporated as a covariate in any models given that only 5% of the students in the sample were English learners and because it was not a focus of the present study. In sum, preliminary analyses indicated that pretest would need to be incorporated into all posttest analyses but not cohort, grade level, or ELL status.

Our final general mixed model included classroom-level experimental condition (effect coded 1 = RVOC treatment, and –1 = control), student-level pretest (standardized as z scores, within classrooms) and classroom-level aggregate pretest (standardized as z scores, across classrooms), much like a traditional analysis of covariance, as follows.

$$\text{Posttest} = \gamma_{000} + \gamma_{010} * \text{Condition} + \gamma_{100} * \text{Z Student Pretest} + \gamma_{020} * \text{Z Classroom Pretest} + U_{00} + r_0 + e$$

In the model just presented, the fixed effects included the conditional mean posttest across students within classrooms and schools (γ_{000}), the fixed effect of RVOC treatment (effect-coded) γ_{010}, and the fixed effects of pretest within and between classrooms (each standardized), γ_{100} and γ_{020}, respectively. The three random effects estimates included posttest variation between schools, classrooms within schools (between classrooms), and students within classrooms (residual), U_{00}, r_0, and e, respectively. Condition and pretests were explicitly assumed to have fixed, not random, effects on posttest because (a) we would expect identical conditions to be employed in a study replication, and (b) we had no reason to believe that the pretest–posttest relationships would systematically vary across schools, and preliminary analyses showed no substantial pretest slope variance. School and classroom membership, on the other hand, were assumed to have random effects on posttest because we would not expect the same schools or classrooms to participate in a study replication, and further, the intercept-only model estimates indicated substantial variance in posttest due to these factors (estimates given in the Results section).

Classroom Literacy Observations

To answer the second research question, we modeled classroom language arts/literacy block instruction to test (a) whether classroom literacy content instructional minutes exhibited linear change during the course of the study, and (b) whether there were differences between conditions on content time, in part to evaluate fidelity and potential contamination (i.e., we expected differences on treatment-related content but not on other content) and in part to assess how RVOC treatment teachers' language arts/literacy content allocations differed from their control teacher peers (i.e., due to being in the treatment condition). For these three-level hierarchical models (measurement occasion nested within classrooms, within schools), time was coded in months commensurate with the observed spacing of the observation periods (0 for baseline, 1.5 since baseline, and 2.7 months since baseline, respectively, for fall, winter, and spring observations; cf. Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). For this set of models, we first estimated mean change across both groups (Model 1), and then we tested RVOC treatment effects (condition covariate was effect coded) on the intercept and slope (Model 2). Note that preliminary analyses revealed no cohort differences (using two effect-coded variables at the classroom

Level 2) on instructional minutes for each content area, and as such, cohort was not included in the final models. For all models, full information maximum likelihood estimation in HLM7 software was employed. Further details about these models can be obtained from the authors upon request.

RESULTS

Rich Vocabulary Effects on Student Outcomes

Descriptive Statistics

Disaggregated student assessment reliabilities, means, and standard deviations for pretests, posttests, and gains, as well as final model-adjusted means for each condition, are provided in Table 6; zero-order correlations for each condition are provided in Appendix A.

Pretest and Posttest Intraclass Correlations

Intercept-only models were estimated first in order to determine the nature of the nonindependence of the data at each level. Results from these simple models (Table 7) show that the median school intraclass correlation (ICC) across outcomes was .06 at pretest and .04 at posttest, and the median classroom intraclass correlation was .14 at pretest and .15 at posttest; combined, 20% or more of the variance in student outcomes was at the classroom and school levels.

Student Pretest Differences

As noted earlier in the Analytic Approach section, preliminary analyses revealed that RVOC treatment students were significantly lower than controls on all pretests, despite our use of random assignment for each cohort of teachers. Specifically, the model-estimated results showed that control students were higher than students assigned to RVOC treatment by an average 1.14 points at pretest (range = 0.86–1.47), which translates to an average difference of 0.22 standard deviations. Unfortunately, randomization does not guarantee that groups will be equal on all characteristics, and in our case they were not. In any simple randomization procedure, the possibility of having significantly different groups by chance is equal to the alpha level (5%). Further, the potential for group nonequivalence to occur by chance in a multicohort, cluster randomized design such as ours is greater because we are dealing with a smaller number of units to be randomized (teachers, not students), and further, our units for randomization were even smaller due to the multicohort nature of the study (with 21, 27, and 14 teachers for each cohort, respectively). Indeed, the smaller the number of units to be assigned, the greater the chance that groups will have any nonzero difference after being randomized, even if the magnitude of the difference is not statistically significant for that sample (Hsu, 1989). These nonzero differences are not detectable for each cohort; however, with the combined cohort data we were able to detect even very small pretest differences of .20 standard deviations. To minimize the selection bias introduced by these nonequivalent groups, pretest was incorporated in all student posttest models, at both the student and the classroom levels, as described in the prior section's Analytic Plan.

Table 6. Observed student assessment descriptives

			RVOC Treatment[a]							Control[b]						
	Cronbach's α		Pretest		Posttest		Gain			Pretest		Posttest		Gain		
Measure	Pre	Post	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*Adj Post M*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*Adj Post M*
Vocabulary																
Proximal																
CBM-Words (max = 28)	.76	.88	15.14	(4.46)	23.60	(4.44)	8.49	(3.97)	23.86	15.82	(4.79)	17.40	(4.92)	1.58	(2.97)	16.88
CBM-Sentences (max = 28)	.56	.71	19.72	(3.31)	23.72	(3.18)	4.03	(3.11)	23.87	20.23	(3.41)	20.99	(3.29)	0.79	(3.01)	20.69
Distal																
ITBS-Vocab (max = 37)	.89	.90	24.60	(7.44)	26.91	(7.22)	2.31	(3.65)	27.13	25.53	(7.41)	27.35	(7.42)	1.81	(3.45)	26.68
Comprehension																
Proximal																
CBM-Compreh (max = 12)	.67	.70	7.61	(2.63)	8.53	(2.64)	0.94	(2.21)	8.73	8.19	(2.43)	8.51	(2.52)	0.32	(1.98)	8.18
Distal																
ITBS-Compreh (max = 43)	.89	.90	29.56	(8.31)	32.07	(8.17)	2.53	(4.17)	32.51	30.89	(7.70)	32.83	(7.84)	1.93	(4.16)	31.87

Note. Raw scores used for all measures; same forms of measures used for pretest and posttest. Cronbach's α = median of internal consistencies across cohorts; RVOC = Rich Vocabulary; Adj Post *M* = adjusted posttest mean from mixed model results estimates (adjusted for dependence within schools and classrooms, as well as student-level and classroom-level pretests); ITBS = Iowa Test of Basic Skills; CBM = curriculum-based measure; Vocab = vocabulary; Compreh = comprehension.

[a]$n = 627$. [b]$n = 605$.

Table 7. Intercept-only (unconditional) three-level random effects models for student assessments at pretest and posttest

	Vocabulary						Comprehension			
	CBM-Words (Proximal)		CBM-Sentences (Proximal)		ITBS-Vocab (Distal)		CBM-Compreh (Proximal)		ITBS-Compreh (Distal)	
Source	*Var*	*ICC*	*Var*	*ICC*	*Var*	*ICC*	*Var*	*ICC*	*Var*	*ICC*
Pretest										
Schools	0.74	.03	0.80***	.07	3.15**	.06	0.17	.03	4.11**	.06
Classrooms	2.98***	.14	0.92***	.08	8.31***	.15	0.81***	.12	9.25***	.14
Residual	17.91		9.67		44.36		5.57		51.77	
Posttest										
Schools	1.20	.04	0.49	.04	4.31***	.08	0.30**	.04	4.96***	.08
Classrooms	14.98***	.46	3.20***	.26	7.21***	.13	0.68***	.10	9.77***	.15
Residual	16.23		8.75		42.45		5.74		50.34	

Note. $N = 1{,}232$ students within 61 classrooms and 24 schools. Raw scores used. Full maximum likelihood estimates reported. CBM = curriculum-based measure; ITBS = Iowa Test of Basic Skills; Vocab = vocabulary; Compreh = comprehension; Var = variance; ICC = intraclass correlation.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Student Posttest Model Results

Results for both vocabulary and comprehension outcomes (shown in Table 8) showed that, adjusted for student and classroom pretest levels, students in RVOC treatment classrooms outperformed their peers in control classrooms on all measures. The estimated adjusted differences between groups on proximal measures were 6.98 points for CBM-Words, 3.18 points for CBM-Sentences, and 0.54 points for CBM-Comprehension. For distal measures, the adjusted group mean differences were estimated at 0.46 and 0.64 points for vocabulary and comprehension, respectively. Because the measures were scaled differently, ESs were calculated simply as an approximate Cohen's *d* (for Condition, computed as twice the coefficient [due to effect coding used], divided by the square root of the sum of the variance components). As can be seen in Table 8, RVOC treatment effects were far larger for the proximal vocabulary outcomes (groups were estimated to average over 1 *SD* apart) compared with the distal measures (approximately 0.15 *SD* apart). ESs were also computed similarly for the effect of individual and classroom aggregate pretest levels on posttest outcomes, except that the coefficient was not doubled. Irrespective of condition, students who were 1 *SD* higher on pretest relative to their classroom peers averaged over 1.19 *SD* higher on posttest, holding all else constant. In addition, students who were in classrooms that had higher pretest skill levels relative to other classrooms in the sample averaged 0.95 *SD* higher on posttest.

One final note is that, as a check on assumption of homogeneity of regression slopes for the pretest covariates, we tested whether student and classroom aggregate pretest levels interacted with experimental condition (both interactions added to the previous model). For brevity, we describe only the salient findings of these models here (additional details regarding model results can be obtained from the authors upon request). Specifically, we found two significant interactions on vocabulary outcomes: one Student Pretest ×

Table 8. Multilevel model results for student outcomes

	Vocabulary						Comprehension			
	CBM-Words (Proximal)		CBM-Sentences (Proximal)		ITBS-Vocab (Distal)		CBM-Compreh (Proximal)		ITBS-Compreh (Distal)	
Fixed Effects	*Coeff*	*ES*	*Coeff*	*ES*	*Coeff*	*ES*	*Coeff*	*ES*	*Coeff*	*ES*
Posttest *M* (Conditional)	20.37***		22.28***		26.90***		8.45***		32.19***	
Condition (Level 2)	3.49***	2.20	1.59***	1.23	0.23*	0.13	0.27***	0.29	0.32*	0.16
Student pretest (Level 1)	3.01***	0.95	1.67***	0.65	6.24***	1.84	1.59***	0.84	6.61***	1.66
Classroom pretest (Level 2)	2.35***	0.74	1.33***	0.52	3.67***	1.08	1.08***	0.57	4.16***	1.05
Random effects	*Var*		*Var*		*Var*		*Var*		*Var*	
Schools	<.01		<.01		0.11		0.04		0.11	
Classrooms	1.40***		0.31***		0.23*		0.02		0.26*	
Residual	8.69		6.33		11.15		3.56		15.42	

Note. N = 1,232 students (627 Rich Vocabulary [RVOC] treatment, 605 control) within 61 classrooms (30 RVOC treatment, 31 control) and 24 schools. Raw scores used for all measures; measure forms were same for pretest and posttest. Condition is effect coded (+1 = RVOC treatment, –1 = control) and tested at classroom level; Student Pretest is group-mean centered and standardized (z score) as number of standard deviations student is from average in his or her classroom, tested at student level; Classroom Pretest are student scores aggregated by classroom and standardized (z scores), tested at classroom level. Full maximum likelihood estimates reported. *ES* = effect size in approximate standard deviation units: for Condition, twice the coefficient divided by the square root of the summed variance components (random effects); for Pretests, coefficient divided by the square root of the summed random effects (ES for Condition can be interpreted as the difference between group posttest means, adjusted for pretest levels, in standard deviations; ES for Pretests can be interpreted as the expected change in posttest for every standard deviation increase in pretest level). CBM = curriculum-based measure; ITBS = Iowa Test of Basic Skills; Vocab = vocabulary; Compreh = comprehension; Var = variance.
*$p < .05$. **$p < .01$. ***$p < .001$.

Condition interaction on the proximal CBM-Words measure (Coeff. = –0.44, $p < .001$), and one Classroom Pretest × Condition interaction on the distal ITBS-Vocabulary measure (Coeff. = –0.22, $p < .05$). Of importance, these two interactions did not change the RVOC treatment effect's significance or its general substantive size, and examination of the model-implied predicted values showed that treatment effects were only slightly smaller if pretest levels were relatively high (small, ordinal interactions). On the proximal vocabulary measure, students who were lower than their own classroom peers at pretest were 7.85 points higher if they were in an RVOC treatment classroom compared to control, whereas students who were higher than classroom peers at pretest were only 6.08 points higher

on posttest if they were in an RVOC treatment classroom compared to control. On the distal vocabulary measure, students in classrooms with relatively lower aggregate pretest levels were 0.92 points higher on posttest if they were in an RVOC treatment classroom compared to control, but treatment students in relatively lower classrooms were only 0.02 points higher than controls by posttest.

RVOC Effects on Classroom Instruction Minutes

As referenced previously in the Methods section, three-level growth models testing for differences between treatment and control classrooms were conducted on the teacher observation data to determine instructional time allocations. For brevity, we report the main results of these models here. (Details can be obtained from the authors upon request.) Results showed that, during the first (fall) observation, RVOC treatment teachers spent significantly more time than control teachers on vocabulary instruction (16.62 min more) and significantly less time than control classrooms on comprehension instruction (9.82 min less). There were no significant differences between groups on linear change in instruction time. In other words, RVOC treatment classrooms spent significantly more time on vocabulary and less time on comprehension than controls, and this did not change over the course of the study. Analyses of vocabulary content subcodes also showed that RVOC treatment classrooms spent 6.84 min more time at the fall observation on word meanings (treatment estimate was 11.63 min compared to control estimate of 4.79 min) and 18.64 min more on contextualizing words (treatment estimate was 21.52 min compared to control classrooms estimated at 2.88 min). The groups, however, did not differ on word relationships (1.68 min for treatment and 1.29 min for controls), nor did they differ on morphological analysis time (treatment estimated time of 1.27 min and control estimate at 0.34 min). Finally, as already mentioned, RVOC treatment teachers were the only teachers observed spending time using treatment lessons (33.07 min), and, with the exception of one control teacher, the only teachers observed using the novels employed in the treatment instruction (7.15 min compared with 0.30 min).

DISCUSSION

The primary aim of this study was to examine the effects of the RVOC approach using a three-cohort cluster randomized design with fourth and fifth graders in urban elementary schools. We examined proximal effects with two custom vocabulary and one custom comprehension measure, and distal effects with norm-referenced measures of vocabulary and comprehension used in the original studies on RVOC.

After adjusting for student and classroom pretest differences, student posttest models showed that there were large, significant effects on the proximal vocabulary measures, the word definition task (CBM-Words), and the sentence verification task (CBM-Sentences), with ESs of 2.20 and 1.23, respectively. The RVOC treatment instruction also had a small, significant effect (0.29) on the custom comprehension measure (CBM-Comprehension), which required students to demonstrate understanding of a passage written to include taught words. In addition, we found significant effects on both distal outcomes, although these ESs were extremely small: for the norm-referenced ITBS vocabulary the ES was 0.13, and for ITBS comprehension, the ES was 0.16.

This pattern of large effects for proximal and small effects for distal outcomes is similar to the findings for many vocabulary interventions, summarized in Elleman et al.'s (2009) meta-analysis, and also similar to the higher Year 1 findings on proximal measures for fourth graders, and the null effects for distal measures reported by Apthorp et al. (2012) in their evaluation of a similar Tier Two vocabulary intervention. Although the RVOC instruction resulted in successful learning of the taught vocabulary words, as measured in a receptive task requiring recognition of the word meaning as well as a more challenging sentence-verification task requiring understanding of the taught words in sentence contexts, the instruction hardly improved general vocabulary knowledge or general comprehension, as measured by the ITBS. Norm-referenced comprehension measures tap a broad set of skills and knowledge, including test-taking skills, that are not really addressed in vocabulary interventions, even one as intense as RVOC (Pearson et al., 2007). We consider the large proximal effects to validate the features of the RVOC approach. Students successfully learned the words that were taught: on word recognition, treatment students gained approximately 30% more words correct above their pretest level, for a total of 84% correct; on word meanings, gained approximately 14% more above their pretest level, for a total of 84% correct; and on comprehension, gained 8% above their pretest level, for a total of 71% correct. Comparatively, controls gained only 6% and 3% beyond pretest levels on the two vocabulary measures and 3% more than pretest on the comprehension measure, totaling 63%, 75%, and 71% on the three measures, respectively. By Grades 4 and 5, written language demands increase and there is value in effectively teaching words that characterize a literate vocabulary. That the distal effects are very small does not seem surprising considering the long-term developmental trajectories of both vocabulary and comprehension skills. The intervention did not explicitly teach general vocabulary or comprehension skills that might improve performance on standardized measures. This does not seem to diminish the educational meaningfulness of the findings. A serious effort to build vocabulary or comprehension seems likely to require sustained and wide use of effective practices like RVOC. Longitudinal gene–environment influences on reading achievement (Olson, Keenan, Byrne, & Samuelsson, 2014) and the complex cognitive, linguistic, and knowledge influences on reading comprehension seem to argue against a "quick fix" to raise general language and comprehension within a discrete intervention period like that of this study (Compton, Miller, Elleman, & Steacy, 2014). Our study design allowed us to demonstrate that the time-intensive RVOC approach significantly influenced learning of taught words. A more challenging research questions is whether *sustained, broader,* or *coordinated* (across content areas) use of this type of instruction might improve performance on distal measures of language and comprehension skills that have long-term developmental trajectories with complex influences. We also examined how teachers in the RVOC treatment condition might differ from typical upper-elementary teachers in classroom instruction time allocations. Classroom observations confirmed that teachers in the RVOC treatment group dedicated significantly more time to vocabulary instruction in word meanings and contextualizing words but not word relationships or morphological analysis and significantly less time to comprehension instruction. Even though the control group teachers were observed to spend more time on comprehension than RVOC teachers, there was a small significant treatment effect for ITBS comprehension. Similar to the teachers in the Apthorp et al. (2012) study on this approach, RVOC teachers engaged students in deep-processing activities related to the Tier Two words, and in higher level inference questions related to the two novels. These features of RVOC may account for distal influences on the standardized comprehension measure. Finally, teachers have a fixed amount of time for language arts/literacy instruction, and if teachers increase their time on vocabulary, they must

necessarily decrease the time on other content. At these grade levels, the decrease would most likely be made in comprehension. The findings from this study on limited comprehension outcomes and the time "cost" of implementing intensive vocabulary intervention that requires reducing time spent on other types of language arts instruction, including comprehension, may clarify the trade-offs teachers must always make. When an intensive vocabulary approach like RVOC is used in an effort to improve global comprehension and language skills for lower skilled students, these findings suggest that the necessary extra instruction and compensatory practice does not fit within the typical school day (see Olson et al., 2014).

Limitations

In this study we attempted to address sampling, design, and analysis limitations of the original studies on rich vocabulary instruction that were conducted in the 1980s.Yet limitations remain in this study. We were unable to recruit a sample that matched the low socioeconomic status and reading levels observed in the earlier studies, and our measures only partially replicate the original measures. Although we report reliabilities for the experimenter-developed measures in this study (data that are often missing in many published vocabulary interventions) the internal consistency levels of our measures at pretest were particularly low for the vocabulary CBM-Sentences and the comprehension CBM. The use of only one passage for CBM comprehension further narrows the skills being tested and reduces confidence in our findings for this construct. These low reliabilities indicate that our ESs for these two measures may be lower than would be expected if they had been more precise measurements. This said, low reliability estimates have also been reported for custom measures used in other vocabulary intervention studies (see Apthorp et al., 2012). Clearly the use of proximal comprehension measures that (a) have been piloted a priori to assure better psychometric properties and (b) allow for assessing a wider range of background knowledge on multiple passages with differing knowledge demands is recommended for future vocabulary studies.

A third limitation is that the implementation data on daily use of the intervention are teacher-self reported. However, the classroom observations confirmed that treatment teachers spent the prescribed amounts of time on vocabulary compared to controls (consistent with teacher self-report log data), and further, that there was no treatment diffusion to control teachers. A stronger test of RVOC would include an alternate vocabulary treatment of similar duration.

A fourth limitation was that, although random assignment was used for each of the three cohorts, controls were found to be significantly higher on pretests. Unfortunately, there is always the possibility that randomization can still lead to group nonequivalence on some baseline characteristic; in any simple randomized trial, the possibility of having significantly different groups by chance is equal to the alpha level. More important, the potential for group nonequivalence to occur by chance in a multicohort, cluster randomized design such as ours is greater because we are necessity dealing with a smaller number of units to be randomized (teachers), particularly given that we randomized separately for each cohort (with 21, 27, and 14 teachers for each cohort, respectively). The smaller the number of units to be assigned, the greater the chance that the groups will have any nonzero difference on a given covariate, even if the magnitude of the difference is not statistically significant (cf. Hsu, 1989; Strube, 1991). Had we analyzed each cohort separately (as if they were three separate studies), the differences between groups on pretest would not be

declared significant; but combined, the difference was detected due to increased power. Future research using cluster randomized trial designs may find this study a case for employing one or two larger cohorts rather than three or more smaller cohorts to avoid the increased possibility of group nonequivalence due to chance alone. All this said, in response to the pretest differences found, we controlled for pretest in all student posttest analyses, both at the individual and the classroom aggregate levels. Because pretests were the exact same forms as the posttests, the selection bias confounding the treatment effect on posttest was minimized to the extent possible. Despite the control group's advantage at pretest, the adjusted treatment effects we found were positive. Finally, we recruited typically developing students in Grades 4 and 5 and provided intervention for 14 weeks to replicate the original studies' conditions. It may be, however, that the treatment would have had greater distal vocabulary or comprehension effects with a different aged sample, a lower skilled sample, or in a small-group implementation (Connor et al., 2014). Again, it may also be that a 14-week whole-class implementation does not allow the targeted instruction tailored to student learning needs to produce generalized vocabulary growth or comprehension benefits. In other words, even this intensive, 14-week, 30-min-a-day classroom treatment that features many evidence-based instructional activities may not offer sufficient individualized opportunities for students to use the words in meaningful contexts to expand broader vocabulary and comprehension skills.

Implications for Instruction and Research

We were left with questions about the impact of RVOC instruction on the teachers, who shared positive feedback about the instruction to researchers both during and after the instructional period. They indicated the value of using prepared and scripted lessons and student materials. Substantial researcher time was involved in designing the instruction based on the two novels used for this study. The team of researchers spent several months developing questions, activities, and prompts that stimulate students to make relations between words and to deeply and actively engage in using the words to discuss each high interest novel. The instruction and materials were quite effective in helping students learn these Tier Two words well. Yet we doubt if most teachers in these cohorts will find the time to prepare similar instruction for other and new novels they teach. We do wonder how the experience of using the materials, teaching these words to develop a literate vocabulary, and teaching vocabulary to a depth of meaning and ownership will influence future vocabulary instruction in these teachers' classrooms. Will the intensive and successful experience with well-designed rich vocabulary materials influence these teachers to devote more time to vocabulary instruction, both during the language arts/literacy block, and in other parts of the school day? Were there specific vocabulary activities that they continue to use and adapt to new instructional settings?

Questions remain about how to view the small distal effects found here and for many vocabulary interventions. We regard the large short-term proximal effects as validation of the RVOC teaching strategies. Yet we do not know how the 14-week RVOC instruction influenced students' use of these more challenging words and their continued vocabulary learning. Anecdotal reports from teachers and our own observations of lively classroom discussions indicated that students were engaged and enjoyed using the words in the contexts provided in the high-interest novels. Researchers carefully designed prompts and questions to appeal to middle school students and to stimulate questions and discussion. Future research might address whether participation in this type of intense academic vocabulary instruction designed to promote depth of understanding influences students' longer term

vocabulary growth or their ability to use these academic words as "tools for learning" (Nagy & Townsend, 2012). Vocabulary size reflects long-term habits and exposures to instruction, reading practice, and word use in speaking and writing. Although distal effects may be insignificant in the short term, sustained use of these approaches may lead to steady accumulation of general vocabulary knowledge, which seems unlikely to occur in relatively discrete interventions like this one. This suggests the value of much longer term research on sustained vocabulary fertilization that optimizes promising features like multiple exposures, review, contextualizing, and building semantic networks across content and grades (e.g., Snow, Lawrence, & White, 2009). Technology may suggest feasible methods for this type of study. Finally, because vocabulary interventions similar to RVOC have had strong effects on learning taught words, it may be more fruitful to examine retention of taught word meanings and their understanding and use in broader school contexts and content areas than to perseverate on the transfer to general vocabulary knowledge and comprehension. That is, the research is clear that vocabulary knowledge has a strong relationship to reading comprehension but that even well-designed vocabulary interventions do not have a direct impact on comprehension. Significant instructional time is required for word learning, and future research might examine the costs of diverting time for vocabulary instruction, and the benefits of this instruction beyond reading comprehension. These benefits may include the long-term effects of an intensive vocabulary intervention on vocabulary growth. These longer term effects may also include increased student awareness and interest in words that support the necessarily incidental learning of most word meanings through exposure in oral and written contexts.

ACKNOWLEDGMENTS

We are most grateful to the teachers and students who participated. We also wish to acknowledge the valuable contributions and comments of our anonymous reviewers.

FUNDING

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant #R305A100568. The opinions expressed are those of the authors and do not represent the views of the US Department of Education.

REFERENCES

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

Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness*, *5*, 160–185.

Baumann, J. F. (2009). Intensity in vocabulary instruction and effects on reading comprehension. *Topics in Language Disorders*, *29*, 312–328.

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

Beck, I. L., & McKeown, M. G. (2004). *Elements of reading vocabulary: Teacher's guide level A*. Austin, TX: Steck-Vaughn.

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

Beck, I. L., McKeown, M. G., & Kucan, L. (2008). *Creating robust vocabulary: Frequently asked questions and extended examples*. New York, NY: Guilford.

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

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

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

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

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, *9*, 30–52.

Blachowicz, C. (1987). Vocabulary instruction: What goes on in the classroom? *Reading Teacher*, *41*, 132–137.

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

Compton, D. L., Miller, A. C., Elleman, A. M., & Steacy, L. M. (2014). Have we forsaken reading theory in the name of "quick fix" interventions for children with reading disability? *Scientific Studies of Reading*, *18*, 55–73.

Connor, C. M., Spencer, M., Day, S. L., Guiliani, 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*, 762–778. doi: 10.1037/a0035921

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

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*, 934–945.

Dale, E., & O'Rourke, J. (1981). *The living word vocabulary*. Chicago, IL: World Book/Childcraft.

Edmonds, M., & Briggs, K. L. (2003). The Instructional Content Emphasis Instrument: Observations of reading instruction. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 31–52). Baltimore, MA: Brookes.

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

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

Freebody, P., & Anderson, R. C. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, *15*, 19–39.

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

Graves, M. F. (1986). Vocabulary learning and instruction. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 13, pp. 49–89). Washington, DC: American Educational Research Association.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2003). *The Iowa Tests of Basic Skills (Forms A and B, Levels 5–14)*. Itasca, IL: Riverside.

Hsu, L. M. (1989). Random sampling, randomization, and equivalence of contrasted groups in psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, *57*, 131–137.

Institute of Education Sciences. (2003). *Identifying and implementing educational practices supported by rigorous evidence*. Washington, DC: U.S. Department of Education. Retrieved from http://www.ed.gov/rschstat/research/pubs/rigorousevid/index.html

Kamil, M. L. (2004). Vocabulary and comprehension instruction. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 213–234). Baltimore, MD: Brookes.

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology*, *100*, 851–868.

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

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., ... Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Loftus, S. M., Coyne, M. D., McCoach, B., Zipoli, R., & Pullen, P. C. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research and Practice*, *25*, 124–136.

Marulis, L. M., & Neuman, S.B. (2013). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal of Research on Educational Effectiveness*, *6*, 223–262.

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

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

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19*, 303–330.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19–36). Hillsdale, NJ: Erlbaum.

Nagy, W. E., & Hiebert, E. H. (2011). Toward a theory of word selection. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4, pp. 388–404). New York, NY: Routledge.

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 269–284). Mahwah, NJ: Erlbaum.

Nagy, W. E., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *47*, 91–108.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston, MA: Heinle.

National Center for Education Statistics. (2012). *The nation's report card: Vocabulary results from the 2009 and 2010 NAEP reading assessments* (NCES 2013 452). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4769). Washington, DC: National Institute of Child Health and Human Development.

Olson, R. K., Keenan, J. M., Byrne, B., & Samuelsson, S. (2014). Why do children differ in their development of reading and related skills? *Scientific Studies of Reading*, *18*, 38–54.

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

Peck, R. (1998). *A long way from Chicago*. New York, NY: Penguin Puffin.

Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.

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

Pressley, M., Disney, L., & Anderson, K. (2007). Landmark vocabulary instructional research and the vocabulary instructional research that makes sense now. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 233–248). New York, NY: Guilford.

Proctor, C. P., Dalton, B., Uccelli, P., Biancarosa, G., Mo, E., Snow, C., & Neugebauer, S. (2011). Improving comprehension online: Effects of deep vocabulary instruction with bilingual and monolingual fifth graders. *Reading and Writing: An Interdisciplinary Journal*, *24*, 517–544.

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

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R & D Program in Reading Comprehension*. Santa Monica, CA: RAND Corporation.

Raymer, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*(Suppl.), 31–74.

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, *83*, 508–525.

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

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, *2*, 325–344.

Snow, C. E., Tabors, P. O., Nicholson, P. A., & Kurland, B. F. (1995). SHELL: Oral language and early literacy skills in kindergarten and first-grade children. *Journal of Research in Childhood Education*, *10*, 37–47.

Spinelli, J. (1990). *Maniac Magee*. New York, NY: Little, Brown.

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

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

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

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

Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist*, *38*, 878–893.

Strube, M. J. (1991). Small sample failure of random assignment: A further examination. *Journal of Consulting and Clinical Psychology*, *59*, 346–350.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Apple Valley, MN: Touchstone Applied Science Associates.

APPENDIX A

Table A1. Rich Vocabulary Instruction word list pool ($N = 140$) sorted by novel, instruction week, and word

Novel Title	Word	Instr Week	CBM Measure (W/S/C)	Freq Split	SFI	U
A Long Way from Chicago	apt	1		high	49.4	8.00
A Long Way from Chicago	codger	1	C	low	22.1	0.02
A Long Way from Chicago	fleeting	1	S	high	40.7	1.00
A Long Way from Chicago	gloat	1	W	low	30.7	0.12
A Long Way from Chicago	grudge	1	S	low	38.3	0.68
A Long Way from Chicago	inquiry	1	W	high	48.9	7.00
A Long Way from Chicago	oblige	1	C	high	38.8	0.76
A Long Way from Chicago	peculiar	1		high	51.8	14.00
A Long Way from Chicago	philanthropist	1	C	low	27.4	0.06
A Long Way from Chicago	savor	1		low	35.0	0.32
A Long Way from Chicago	capable	2		high	54.4	27.00
A Long Way from Chicago	culprit	2	W	high	39.5	0.88
A Long Way from Chicago	glower	2	S	low	29.3	0.09
A Long Way from Chicago	hearty	2		high	42.4	1.00
A Long Way from Chicago	jaunt	2	C	low	27.7	0.06
A Long Way from Chicago	lament	2	W	low	39.2	0.84
A Long Way from Chicago	pilfer	2	C	low	27.7	0.06
A Long Way from Chicago	sanitary	2		high	45.3	3.00
A Long Way from Chicago	tragic	2	S	high	48.5	7.00
A Long Way from Chicago	waver	2		low	36.0	0.40
A Long Way from Chicago	brisk	3		high	46.3	4.00
A Long Way from Chicago	fascinated	3	S	high	49.7	9.00
A Long Way from Chicago	linger	3	C	high	41.5	1.00
A Long Way from Chicago	loom	3		high	48.6	7.00
A Long Way from Chicago	prim	3	C	low	34.1	0.25
A Long Way from Chicago	relentless	3	W	high	41.9	1.00
A Long Way from Chicago	spry	3		low	36.4	0.44
A Long Way from Chicago	stagnant	3	S	low	39.9	0.99
A Long Way from Chicago	stifle	3	C	low	37.5	0.56
A Long Way from Chicago	writhe	3	W	low	31.1	0.13
A Long Way from Chicago	dainty	4	C	low	42.4	1.00
A Long Way from Chicago	frantic	4	W	high	45.7	3.00
A Long Way from Chicago	maneuver	4	S	high	43.5	2.00
A Long Way from Chicago	novelty	4		low	43.4	2.00
A Long Way from Chicago	reconcile	4	S	low	40.9	1.00
A Long Way from Chicago	revel	4	W	low	33.0	0.20
A Long Way from Chicago	stern	4		high	49.5	8.00
A Long Way from Chicago	tension	4		high	51.7	14.00
A Long Way from Chicago	torment	4	C	high	43.5	2.00
A Long Way from Chicago	transfixed	4		low	34.5	0.28
A Long Way from Chicago	abdicate	5	S	low	34.9	0.31
A Long Way from Chicago	behold	5		high	44.4	2.00
A Long Way from Chicago	bewitched	5		low	37.3	0.53
A Long Way from Chicago	capacity	5	S	high	54.3	26.00
A Long Way from Chicago	confounded	5		low	37.5	0.56
A Long Way from Chicago	defy	5		high	42.1	1.00

(Continued on next page)

Table A1. *(Continued)*

Novel Title	Word	Instr Week	CBM Measure (W/S/C)	Freq Split	SFI	U
A Long Way from Chicago	hunker	5		low	22.1	0.02
A Long Way from Chicago	pry	5	W	high	40.5	1.00
A Long Way from Chicago	savage	5	C	high	49.6	9.00
A Long Way from Chicago	shroud	5	W	low	32.1	0.16
A Long Way from Chicago	bogus	6	C	low	33.9	0.24
A Long Way from Chicago	dignity	6	W	high	50.8	11.00
A Long Way from Chicago	falsify	6		low	30.4	0.11
A Long Way from Chicago	glimmer	6	W	low	41.8	1.00
A Long Way from Chicago	grasp	6		high	51.3	13.00
A Long Way from Chicago	riddled	6	S	low	22.1	0.02
A Long Way from Chicago	sacred	6		high	51.2	13.00
A Long Way from Chicago	taut	6		low	44.6	2.00
A Long Way from Chicago	trait	6		high	50.5	11.00
A Long Way from Chicago	vicinity	6	S	high	45.6	3.00
A Long Way from Chicago	brandish	7		low	22.1	0.02
A Long Way from Chicago	conspire	7	S	low	32.4	0.17
A Long Way from Chicago	gallivanting	7		low	26.9	0.05
A Long Way from Chicago	marvel	7	W	high	43.4	2.00
A Long Way from Chicago	mire	7	W	low	36.1	0.41
A Long Way from Chicago	primed	7	S	high	38.7	0.75
A Long Way from Chicago	restrain	7		high	41.9	1.00
A Long Way from Chicago	stature	7	C	high	44.7	2.00
A Long Way from Chicago	vermin	7		low	38.0	0.63
A Long Way from Chicago	weary	7	C	high	48.7	7.00
Maniac Magee	accurate	1		high	55.4	34.00
Maniac Magee	blunder	1	W	low	36.2	0.41
Maniac Magee	clamor	1	S	high	42.2	1.00
Maniac Magee	glare	1	W	high	47.6	5.00
Maniac Magee	infamous	1	C	low	37.9	0.62
Maniac Magee	jolt	1	C	low	40.5	1.00
Maniac Magee	legacy	1	C	high	40.7	1.00
Maniac Magee	lunge	1		low	37.9	0.45
Maniac Magee	pandemonium	1	S	low	33.7	0.23
Maniac Magee	theory	1	C	high	59.4	86.00
Maniac Magee	befuddled	2	C	low	25.1	0.03
Maniac Magee	blemish	2	W	low	33.5	0.22
Maniac Magee	converge	2		low	40.0	0.99
Maniac Magee	cringe	2	C	low	31.7	0.15
Maniac Magee	instinct	2		high	50.0	9.00
Maniac Magee	mend	2		high	45.1	3.00
Maniac Magee	prospect	2		high	48.7	7.00
Maniac Magee	solemn	2	W	high	47.9	6.00
Maniac Magee	solitude	2	S	low	43.2	2.00
Maniac Magee	stunned	2	S	high	47.2	5.00
Maniac Magee	alert	3	S	high	51.9	15.00
Maniac Magee	cunning	3	C	high	46.4	4.00
Maniac Magee	escort	3	W	low	42.7	1.00
Maniac Magee	flinch	3	S	low	35.6	0.37

(Continued on next page)

Table A1. *(Continued)*

Novel Title	Word	Instr Week	CBM Measure (W/S/C)	Freq Split	SFI	U
Maniac Magee	grim	3		high	48.6	7.00
Maniac Magee	opponent	3	W	high	47.5	5.00
Maniac Magee	publicize	3	C	low	37.5	0.56
Maniac Magee	regret	3		high	47.1	5.00
Maniac Magee	spectator	3		low	41.3	1.00
Maniac Magee	testimony	3		low	45.5	3.00
Maniac Magee	climax	4		high	47.3	5.00
Maniac Magee	dumbfounded	4	C	low	36.4	0.44
Maniac Magee	flaunt	4	C	low	25.3	0.03
Maniac Magee	frayed	4	W	high	41.5	1.00
Maniac Magee	notion	4	S	high	51.4	13.00
Maniac Magee	preposterous	4	S	low	36.1	0.41
Maniac Magee	proclaim	4		high	41.1	1.00
Maniac Magee	prod	4		low	36.3	0.43
Maniac Magee	robust	4	C	high	40.3	0.60
Maniac Magee	sleazy	4	W	low	30.0	0.10
Maniac Magee	careen	5	W	low	22.1	0.02
Maniac Magee	disperse	5		low	41.1	1.00
Maniac Magee	meander	5	S	low	29.3	0.09
Maniac Magee	retaliate	5		low	37.9	0.62
Maniac Magee	stoic	5		low	34.8	0.30
Maniac Magee	timid	5		high	46.6	4.00
Maniac Magee	vague	5	W	high	50.3	10.00
Maniac Magee	vast	5	S	high	57.0	50.00
Maniac Magee	venture	5		high	48.2	6.00
Maniac Magee	wary	5		high	43.1	2.00
Maniac Magee	agony	6		high	47.3	5.00
Maniac Magee	devour	6		high	42.4	1.00
Maniac Magee	ecstatic	6	S	high	36.6	0.46
Maniac Magee	extort	6	C	low	27.4	0.06
Maniac Magee	exuberant	6	W	high	41.4	1.00
Maniac Magee	ludicrous	6		low	35.6	0.36
Maniac Magee	nonchalant	6	S	low	31.6	0.15
Maniac Magee	perilous	6		high	41.8	1.00
Maniac Magee	reprisal	6	W	low	27.4	0.06
Maniac Magee	zany	6	C	low	23.0	0.02
Maniac Magee	diverge	7	W	low	34.9	0.31
Maniac Magee	gauntlet	7		low	34.8	0.30
Maniac Magee	ignorant	7	S	high	48.3	6.00
Maniac Magee	illusion	7		high	46.3	4.00
Maniac Magee	legendary	7		high	44.9	2.00
Maniac Magee	merciful	7	S	low	39.0	0.79
Maniac Magee	persist	7		high	44.8	3.00
Maniac Magee	saunter	7	C	low	32.5	0.18
Maniac Magee	Shudder	7	W	high	41.8	1.00
Maniac Magee	Veer	7		low	34.9	0.31

Note. CBM = curriculum-based measures; W = CBM-Words; S = CBM-Sentences; C = CBM-Comprehension; SFI = Standard Frequency Index; U = frequency of word type per million tokens.

Table A2. Zero-order (disaggregated) correlations among outcome variables and RVOC treatment-specific content

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
Classroom Obs Mean Minutes												
1. RVOC Treatment Lesson Use	—	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2. RVOC Treatment Text Use	—	—	−.07	−.05	−.04	−.08*	−.07	−.07	−.06	−.02	−.07	−.08
Student Pretests												
3. CBM-Words (Proximal)	−.01	.09*	—	.61***	.76***	.56***	.74***	.81***	.64***	.76***	.59***	.69***
4. CBM-Sentences (Proximal)	.02	.14***	.59***	—	.62***	.48***	.60***	.60***	.61***	.62***	.49***	.58***
5. ITBS- Vocab (Distal)	−.02	.12**	.80***	.60***	—	.64***	.81***	.78***	.65***	.89***	.67***	.79***
6. CBM-Compreh (Proximal)	−.01	.08*	.58***	.44***	.68***	—	.72***	.56***	.48***	.66***	.68***	.66***
7. ITBS-Compreh (Distal)	−.07	.11**	.69***	.56***	.83***	.72***	—	.74***	.63***	.82***	.71***	.86***
Student Posttests												
8. CBM-Words (Proximal)	.01	.12**	.61***	.51***	.73***	.59***	.69***	—	.66***	.75***	.59***	.72***
9. CBM-Sentences (Proximal)	−.01	.09*	.59***	.55***	.67***	.56***	.64***	.74***	—	.70***	.51***	.63***
10. ITBS-Vocab (Distal)	−.03	.14***	.72***	.62***	.88***	.70***	.81***	.79***	.65***	—	.71***	.83***
11. CBM-Compreh (Proximal)	−.03	.11**	.52***	.45***	.65***	.65***	.71***	.63***	.58***	.71***	—	.78***
12. ITBS-Compreh (Distal)	.00	.08*	.67***	.55***	.82***	.71***	.87***	.74***	.68***	.84***	.76***	—

Note. Rich Vocabulary (RVOC) treatment condition in lower diagonal ($N = 627$ students within 30 classrooms) and control condition in upper diagonal ($N = 605$ students within 30 classrooms). Classroom Obs Mean Minutes = mean language arts/literacy block minutes per day on treatment-specific lesson and text use across three observations (fall, winter, spring); raw scores used for all student pretests and posttests; same forms used for pretest and posttest; CBM = curriculum-based measure; ITBS = Iowa Test of Basic Skills; Vocab = vocabulary; Compreh = comprehension.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

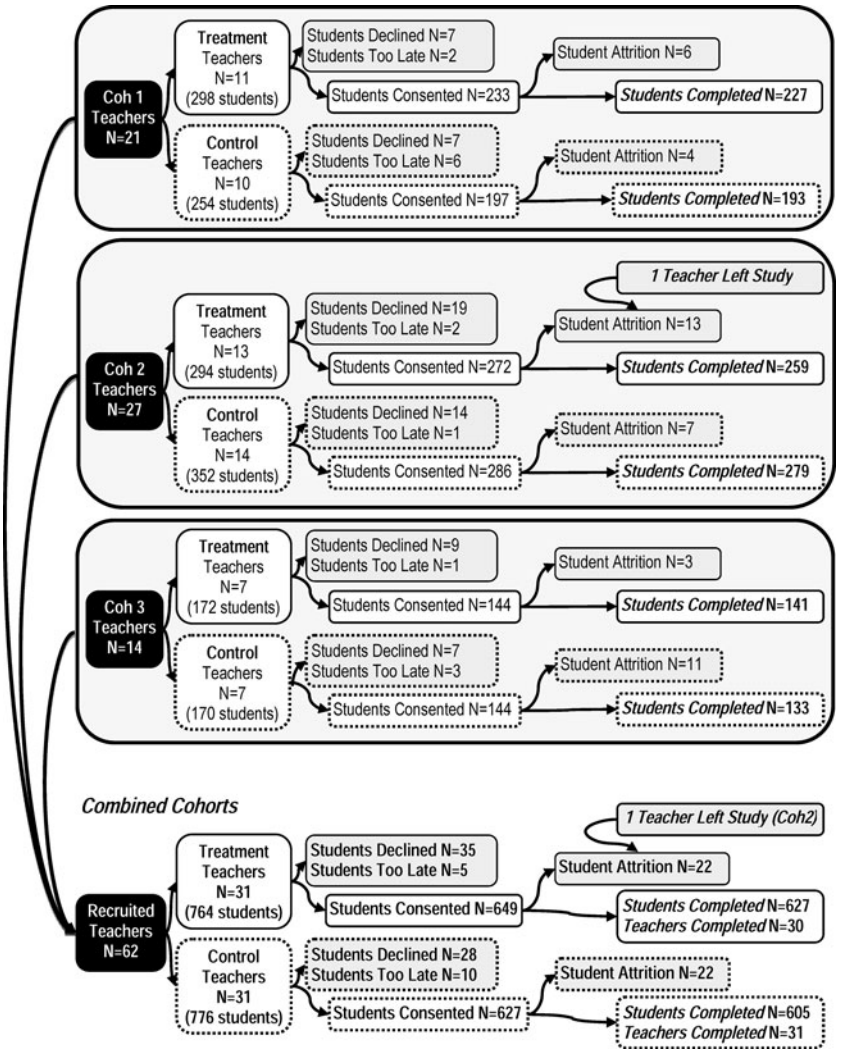

Figure A1. Flowchart of recruitment, assignment, and attrition by cohort (Coh). *Note*. Treatment = Rich Vocabulary instruction.

APPENDIX B

Directions:

Each question below has an unfinished sentence with an underlined word. The question and answer choice will be read aloud as you follow along. Circle the answer below the sentence that is closest in meaning to the underlined word.

Example:

Here is an example: **A furious person is:**

a. excited
b. angry
c. confused
d. annoyed

Answer: If you are a furious person, you are angry, so you would circle **b** for angry.

Reminders:

- Follow along as each question is read aloud.
- Complete each question by circling the letter next to the word that is closest in meaning to the underlined word.
- Please keep your answers to yourself. Don't say them out loud, just mark your answer and then cover it with your colored paper.
- If you are unsure, make your best guess so that you have one answer circled for each question.

Questions:

1. **An escort is a person who:**
   a. cleans houses
   b. goes with you
   c. has many friends
   d. is in jail

2. **To marvel is to:**
   a. show surprise
   b. annoy someone
   c. tell a secret
   d. set something free

3. **To gloat is to:**
   a. borrow
   b. grab
   c. steer
   d. brag

4. **A glimmer is a:**
   a. sharp object
   b. small sign
   c. light rain
   d. narrow opening

5. **A blunder is a:**
   a. painful injury
   b. careless mistake
   c. historic event
   d. difficult task

6. **Something frayed is:**
   a. sealed shut
   b. hard to see
   c. brand new
   d. coming apart

7. **An inquiry is a:**
   a. search for information
   b. false statement
   c. large reward
   d. strongly held belief

8. **To lament is to:**
   a. quickly forget
   b. choose by vote
   c. show sadness
   d. try to understand

9. **To glare is to:**
   a. use all your strength
   b. give an angry look
   c. take completely apart
   d. leave in a hurry

Go on …

Figure B1. CBM-Words.

10. A mire is a:
a. musical instrument
b. loud noise
c. deep cave
d. bad situation

11. A solemn person is:
a. constantly moving
b. very serious
c. painfully shy
d. overly proud

12. An opponent is a person who:
a. builds houses
b. has no family
c. lives far away
d. is a competitor

13. To revel is to:
a. enjoy
b. be sorry
c. dig up
d. go back

14. Something relentless is:
a. obviously fake
b. never ending
c. very deep
d. slightly crooked

15. A reprisal is something you do to:
a. make maps
b. get revenge
c. dig tunnels
d. explain an idea

16. A blemish is a:
a. long spear
b. difficult goal
c. happy memory
d. small mark

17. An exuberant person is:
a. always late
b. full of energy
c. well dressed
d. very bossy

18. A sleazy place is:
a. dirty
b. hidden
c. organized
d. crowded

19. To shudder is to:
a. stumble over words
b. climb downwards
c. walk clumsily
d. shake with fear

20. To pry is to:
a. work hard
b. lose confidence
c. pull apart
d. follow closely

21. To careen is to:
a. give an opinion
b. stand in front
c. understand completely
d. rush forward wildly

22. To writhe is to:
a. eat in a hurry
b. think carefully
c. twist violently
d. give money

23. A frantic person is:
a. very shy
b. not trustworthy
c. stressed out
d. easy going

24. A person with dignity has:
a. hidden anger
b. self respect
c. physical strength
d. great beauty

25. A culprit is a person who:
a. is guilty
b. lives alone
c. preaches
d. owns a business

26. To shroud something is to:
a. worship it
b. cover it
c. shred it
d. avoid it

27. To diverge is to:
a. make a promise in writing
b. give a convincing speech
c. go in a different direction
d. use again in a new way

28. Something vague is:
a. fragile
b. unclear
c. smelly
d. messy

Figure B1. Continued.

Directions:

You are going to try to figure out whether the questions below should be answered yes or no. The questions will be read aloud as you follow along. Try to imagine what the question says and figure out if it makes sense or not. If it makes sense, circle yes. If it doesn't make sense or seems impossible, circle no.

Examples:

Here is an example: If you imagine something, are you picturing it? You would say **yes** because you usually imagine something by picturing it in your mind. So, imagining and picturing match.

Here is another example: If you imagine something, are you ignoring it? You would say **no**, because imagining something and ignoring something do not match.

Practice Questions:

A. If you are joyful, are you sad? yes no

B. If you frighten someone, did you scare them? yes no

Reminders:

- Follow along as each question is read aloud.
- Complete each question by circling either yes or no.
- Please keep your answers to yourself. Don't say them out loud, just mark your answer and then cover it with your colored paper.
- If you are unsure, make your best guess so that you have one answer circled for each question.

Questions:		Circle One	
1.	If you have a notion, are you thinking?	yes	no
2.	If you are stunned, are you relaxed?	yes	no
3.	If something is fleeting, does it last a long time?	yes	no
4.	If you are primed, are you ready?	yes	no
5.	If you glower, do you laugh loudly?	yes	no
6.	If you are merciful, are you kind?	yes	no
7.	If you are fascinated, are you bored?	yes	no
8.	If you maneuver something, do you leave it alone?	yes	no
9.	If you are ignorant, do you have all of the facts?	yes	no

Go on …

Figure B2. CBM-Sentences.

Questions:		Circle One	
10.	If something is vast, is it large?	yes	no
11.	If you abdicate, do you give up something?	yes	no
12.	If you are nonchalant, are you nervous?	yes	no
13.	If there is solitude, are you in a crowd?	yes	no
14.	If you conspire, do you secretly plan?	yes	no
15.	If you hold a grudge, are you forgiving?	yes	no
16.	If you clamor, are you demanding?	yes	no
17.	If you meander, are you in a hurry?	yes	no
18.	If something is tragic, is it very sad?	yes	no
19.	If something is preposterous, is it hard to believe?	yes	no
20.	If you are riddled with something, do you have a lot of it?	yes	no
21.	If you reconcile with someone, do you fight against them?	yes	no
22.	If something is stagnant, is it changing?	yes	no
23.	If you have the capacity for something, are you prepared for it?	yes	no
24.	If there is pandemonium, are people confused?	yes	no
25.	If a place is in the vicinity, is it far away?	yes	no
26.	If you flinch, do you move suddenly?	yes	no
27.	If you are alert, are you paying attention?	yes	no
28.	If you are ecstatic, are you angry?	yes	no

Total correct: _____

Figure B2. Continued.

Read the story to yourself and then answer the questions on the next page.

Horses for Rent!

Sally begged for a horse for her eighth birthday. To her surprise, her parents agreed! She used her cunning bargaining skills to get a second horse, and then a third, a robust filly named Jim. By the time she'd turned twelve, Sally had twenty-seven horses and four dainty ponies. Sally was so obliged to her parents that she wanted to earn some money for her family!

The mayor of Copperville was a codger named Augustus. His legacy was that he had outlawed cars. Sally had a theory that a horse rental business would give the weary townspeople a quicker way to get around. She began to rent her horses to the sauntering folk of Copperville from a prim little stand she'd set up in her parents' front yard. Sure enough, it quickly became fashionable to take a Sunday jaunt on a beautiful stallion.

That is, until a zany boy named Horace opened a motorcycle rental business beside his father's auto shop. Apparently, the befuddled mayor didn't think to outlaw motorcycles! And motorcycles were much faster than horses.

Horace's motorcycles were loud. Their savage roars made many townspeople cringe. This spelled disaster for Sally's business, since the roars tormented the horses too. Riders began to complain of being thrown to the ground and watching their horses jolt off without them. Business was stifled. Finally, on one sunny Sunday afternoon while her horses lingered in their pen, a weary Sally shut her doors completely as her eyes filled with tears.

One day, a savage roar arose out front. Sally thought it was just Horace flaunting the new motorcycle he was able to buy after pilfering her customers. When she peeked through the curtains she saw Horace and his father jolt onto the lawn and saunter up to her door.

Horace's infamous father Zed was a cunning man. He had been run out of nearby Iron City for extorting money from the rich and for selling the poor bogus vacuum cleaners. Sally had no desire to speak to a man of such low stature. But, she couldn't leave them lingering on the doorstep. So she opened the door.

"I've closed up shop," Horace told Sally. "Copperville's had its fill of my zany cycles. I'm sorry, Sally. I never meant to stifle your horse rentals." Sally was befuddled by this. What Zed said next left her dumbfounded.

"We want to help you reopen your business," Zed said. "Give you whatever you need to start a real legacy."

Sally couldn't believe her ears. When had Zed become a philanthropist? One theory was that he was tormented by his bad reputation from his infamous dealings in Iron City. The reason did not matter to Sally. She was much obliged to Zed and the three of them shook hands and smiled and laughed together.

Sally's reopening was publicized in the Copperville Gazette. Her business's stature grew quickly. It remains a robust business to this day, in the heart of the prim, car-free town of Copperville.

Go on ...

Figure B3. CBM-Comprehension.

Circle the one answer you think is the best for each question below about the story *Horses for Rent!*:

1. Why did Sally want to open a horse rental business?
   a. To give back to her family
   b. To exercise all her horses
   c. To help the town have fun
   d. To save money for more horses

2. Why did Sally think a horse rental business would succeed in Copperville?
   a. People love to go on horse rides
   b. There were many great trails for riding
   c. The mayor did not allow cars
   d. Renting horses was very popular

3. How did the townspeople react when Sally opened her horse rental business?
   a. They were glad it attracted more tourists
   b. They were scared of the horses
   c. They realized the horses were useful
   d. They thought it was fun right away

4. What happened when Horace's business opened?
   a. Everyone switched to riding motorcycles
   b. People boycotted Horace's business
   c. Many people were hurt riding motorcycles
   d. Sally could no longer rent her scared horses

5. How did Sally feel when she closed her business?
   a. Depressed that she had failed
   b. Furious about Horace's motorcycles
   c. Relieved to have more free time
   d. Worried about her horses

6. When Sally heard Horace coming, what did she think he wanted to do?
   a. Show off his new motorcycle
   b. Offer her a great new job
   c. Try to become friends
   d. Take what money she had left

7. Why didn't Sally want to speak to Horace's father Zed?
   a. Everyone thought he was a bad man
   b. She was angry at his son
   c. His showing off made her angry
   d. She didn't need a vacuum cleaner

8. What had Horace actually come to tell Sally?
   a. He wanted to take care of her horses
   b. He wanted to combine their businesses
   c. He was closing his own business
   d. He wanted to buy one of Sally's horses

9. What did Zed offer Sally?
   a. To buy her horse business
   b. To help her be successful again
   c. To build her a nicer shop
   d. To sell her a vacuum cleaner

10. How did Sally feel when Zed offered to help?
    a. Suspicious about why he wanted to help
    b. Sorry she had misjudged him
    c. Shocked he was doing a good thing
    d. Satisfied that she finally got her way

11. Some people thought that Zed wanted to help Sally because he wanted to:
    a. Trick her out of her money
    b. Take over her business
    c. Return to Iron City
    d. Improve his bad reputation

12. What happened to Sally's business after it reopened?
    a. It stayed open only a short time
    b. It moved to the center of town
    c. It continues to be a strong business
    d. It was sold to Horace

Figure B3. Continued.