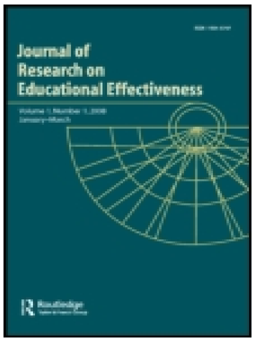

Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

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

Effectiveness of Supplemental Kindergarten Vocabulary Instruction for English Learners: A Randomized Study of Immediate and Longer-Term Effects of Two Approaches

Patricia F. Vadasy, Elizabeth A. Sanders & J. Ron Nelson

To cite this article: Patricia F. Vadasy, Elizabeth A. Sanders & J. Ron Nelson (2015) Effectiveness of Supplemental Kindergarten Vocabulary Instruction for English Learners: A Randomized Study of Immediate and Longer-Term Effects of Two Approaches, Journal of Research on Educational Effectiveness, 8:4, 490-529, DOI: 10.1080/19345747.2015.1033120

To link to this article: https://doi.org/10.1080/19345747.2015.1033120

Accepted author version posted online: 11 Apr 2015.
Published online: 09 Oct 2015.

Journal of Research on Educational Effectiveness, 8: 490–529, 2015

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345747.2015.1033120

Effectiveness of Supplemental Kindergarten Vocabulary Instruction for English Learners: A Randomized Study of Immediate and Longer-Term Effects of Two Approaches

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

Elizabeth A. Sanders
University of Washington, Seattle, Washington, USA

J. Ron Nelson
University of Nebraska–Lincoln, Lincoln, Nebraska, USA

Abstract: A two-cohort cluster-randomized trial was conducted to estimate effects of small-group supplemental vocabulary instruction for at-risk kindergarten English learners (ELs). Connections students received explicit instruction in high-frequency decodable root words, and interactive book reading (IBR) students were taught the same words in a storybook reading context. A total of 324 EL students representing 24 home languages and averaging in the 10th percentile in receptive vocabulary completed the study (Connections $n = 163$ in 75 small groups; IBR $n = 161$ in 72 IBR small groups). Although small groups in both conditions made significant immediate gains across all measures, Connections students made significantly greater gains in reading vocabulary and decoding ($d = .64$ and .45, respectively). At first-grade follow-up, longer-term gains were again greater for Connections students, but with smaller effect sizes ($d = .29$ and .27, respectively). Results indicate that explicit Connections instruction features designed to build semantic, orthographic, and phonological connections for word learning were effective for improving proximal reading vocabulary and general decoding; however, increases in root word reading vocabulary did not transfer to general vocabulary knowledge.

Keywords: English learners, kindergarten, multi-level modeling, supplemental instruction, cluster-randomized trial

The vocabulary knowledge that young children acquire during their preschool and early school years forms a requisite base for future school achievement. Vocabulary knowledge supports early reading development (Roskos et al., 2008), and predicts later reading comprehension (Cromley & Azevedo, 2007; Ricketts, Nation, & Bishop, 2007). The complex nature of vocabulary knowledge prepares students to develop language and strategic skills that lead to greater word learning (e.g., Cunningham & Stanovich, 1993; Storch & Whitehurst, 2002). Gaps in vocabulary knowledge between lower- and higher-skilled students emerge early and widen over time (Biemiller, 2005). Two risk factors are well established. Children from language minority backgrounds who enter school not yet proficient in English are at risk for low levels of English vocabulary (August, Carlo, Dressler, & Snow,

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

2005; Kieffer & Vukovic, 2013). Children of parents with low income and education levels have smaller vocabularies than children of parents with higher income and education (Graves, 2006; Hart & Risley, 1995; Hoff, 2003; Rowe, Raudenbush, & Goldin-Meadow, 2012). Further, these two populations often overlap, sharing environmental influences that create a word knowledge gap that influences later reading (Farkas & Beron, 2004; Kieffer & Vukovic, 2013; Rowe et al., 2012). It is both critical and challenging to begin early to narrow these initial gaps in vocabulary knowledge (Biemiller & Slonim, 2001; Hart & Risley, 2003; Marulis & Neuman, 2013).

ENGLISH VOCABULARY KNOWLEDGE OF ENGLISH LEARNERS

Children who are English learners (EL) by definition have limited English proficiency and English vocabularies (August et al., 2005; Bialystok, Luk, Peets, & Yang, 2010; Swanson, Saez, & Gerber, 2006; Verhoeven, 2011). Research on EL students outlines the cascading effects of limited vocabulary knowledge: kindergarten English vocabulary predicts first-grade word reading and reading fluency (Swanson, Rosston, Gerber, & Solari, 2008; Yesil-Dagli, 2011) and influences later word reading efficiency and comprehension (Kieffer & Vukovic, 2013; Lindsey, Manis, & Bailey, 2003; Proctor, Carlo, August, & Snow, 2005). At Grades 4, 8, and 12, English learners have average English vocabulary scores lower than non-English learners (National Center for Education Statistics, 2012; Skibbe, Connor, Morrison, & Jewkes, 2011), although English vocabulary scores reflect only part of a bilingual child's vocabulary knowledge (Oller & Pearson, 2002).

Schooling alone appears to have little influence on children's vocabulary growth (Christian, Morrison, Frazier, & Massetti, 2000; Skibbe et al., 2011). Studies of vocabulary instruction in K–1 classrooms reveal that teachers devote limited time to teaching vocabulary (Wright, 2012; Wright & Neuman, 2013), and instruction varies widely in how well it reflects research on effective vocabulary practices (Marulis & Neuman, 2010, 2013; Wright, 2012). The need for improving the reading success of English learners has raised calls for supplemental interventions to build their language and English vocabulary knowledge (Kieffer & Vukovic, 2013; Loftus, Coyne, McCoach, Zipoli, & Pullen, 2010).

VOCABULARY APPROACHES FOR YOUNG CHILDREN AT RISK FOR ACADEMIC UNDERACHIEVEMENT

Vocabulary approaches for at-risk preschool and primary-age children include storybook read-alouds and direct instruction. The theoretical basis for the storybook approach holds that the contexts in which words are spoken are the source of most vocabulary learning (see Bolger, Balass, Landen, & Perfetti, 2008). Storybook read-alouds with young children provide rich, engaging, accessible story contexts that may support learning word meanings. Story contexts provide verbal and concrete referents that help children associate new words with their meanings (McKeown & Beck, 2014). Storybook reading is a widely used practice with an extensive research base to support benefits for developing literacy and language skills for young preschool and kindergarten children (Bus, van IJzendoorn, & Pellegrini, 1995; Mol & Bus, 2011; Penno, Wilkinson, & Moore, 2002; Robbins & Ehri, 1994; Whitehurst et al., 1999), and features that enhance vocabulary learning include multiple exposures to new words (Elley, 1989), and explicit explanations of words (Coyne, Simmons, Kame'enui, & Stoolmiller, 2004; Justice, Meier, & Walpole, 2005).

Storybook reading procedures to increase young children's print and alphabet knowledge may also support word learning, including explicit references by adults to printed words during shared reading (Justice & Ezell, 2002; Justice, Kaderavek, Fan, Sofka, & Hunt, 2009). In their review of preschool and kindergarten interactive book reading (IBR) interventions, Mol, Bus, and de Jong (2009) reported a moderate general effect for expressive vocabulary, as well as for alphabet knowledge in kindergarten when teachers more often call attention to print features in storybooks. Directing attention to the written spellings of vocabulary words has been found to support vocabulary learning during read-alouds in kindergarten (Silverman, 2007) and older children (Ricketts, Bishop, & Nation, 2008). Silverman and Crandell (2010) found that teaching the sounds and spellings of target vocabulary was associated with proximal (learning of taught words) receptive vocabulary learning. Of particular note is a treatment comparison by Bowyer-Crane et al. (2007). At-risk 4- to 5-year-olds were randomly assigned to either a reading and phonology (RP) or an oral language (OL) intervention. RP treatment included instruction in word-level decoding reinforced with storybook reading and writing; OL treatment featured vocabulary, grammar, and language skills taught in the context of creating and elaborating on stories. Trained research assistants provided a mix of individual and small-group instruction for 20–30 min/day for 20 weeks. Children in the RP treatment made better progress in phonological awareness and early reading skills, and children in the OL treatment made better progress in vocabulary and grammatical skills. Effect sizes for both treatments were moderate to large and maintained at a 5-month follow-up.

Storybook reading offers a supportive context to help young English learners acquire vocabulary. Roberts and Neal (2004) found that preschool children with beginning levels of English proficiency derived benefit from explicit vocabulary instruction in a story-reading context. Collins (2010) found that a rich explanation approach to teaching vocabulary words in a storybook read-aloud to preschool English learners resulted in high levels of learning for low-frequency target words (e.g., *submerged*, *montage*, *aperture*, *corona*). Instruction included pointing to an illustration of the word, explaining the meaning, providing a synonym, using gestures, and using the word in a context different from that in which it appeared in the story. Children receiving the rich instruction learned about half of the taught target words, further supporting that sophisticated words can be learned by English learners in a storybook context.

Interactive book reading (IBR; Wasik & Bond, 2001; Wasik, Bond, & Hindman, 2006) is a widely used adult–child storybook reading approach designed to help at-risk preschool-aged children develop vocabulary. An adult typically reads a storybook aloud, previewing and introducing words, and asking open-ended questions to elicit discussion of story context. Wasik and Bond (2001) reported significant improvement in knowledge of taught words and general expressive vocabulary, including benefits for children with limited language skills (Wasik et al., 2006). In their meta-analysis of preschool and kindergarten interactive and shared book reading vocabulary interventions, Marulis and Neuman (2010) found that explicit interventions that featured instruction in word meanings and relationships had larger effects ($g = 1.10$) than implicit (0.62), and effect sizes were larger for researcher-developed (1.21) compared to standardized (0.71) measures. However, vocabulary interventions were least effective for lower SES at-risk students, and did not close the gap in vocabulary knowledge between at-risk and average learners. The IBR procedures in this study incorporate features of effective vocabulary instruction identified by other researchers (e.g., Marulis & Neuman, 2010; Silverman & Hines, 2009; Zipoli, Coyne, & McCoach, 2011).

Children also learn new vocabulary through explicit instruction in which the meaning of words are clearly defined or explained, often with multiple exposures and opportunities to interact with the word. Explicit instruction approaches are often combined with storybook reading. Explicit vocabulary instruction that features a dual focus on the word form (spelling and pronunciation) and meaning is guided by lexical-processing models that describe interrelationships among semantic, phonological, and orthographic word features (see Adams, 1990; McClelland, Rumelhart, & PDP Research Group, 1986). As spelling and decoding skills become more accurate, the orthographic and phonological processors form stronger, faster, and more accurate connections with the meaning processor to allow rapid pronunciation, spelling, and word-meaning access. Recent research indicates that even in kindergarten the word-meaning, word-reading, and spelling processes are mutually facilitative (Kim, Al Otaiba, Puranik, Folsom, & Gruelich, 2014). The explicit instruction approach in this study fits this processing model in presenting the word meaning and providing opportunities to decode and spell target words. Explicit instruction that helps children develop accurate and complete memory for phoneme-grapheme mapping and spelling may be helpful for English learners. Research on older English learners indicates the difficulties with orthographic learning (Schwartz, Kahn-Horwitz, & Share, 2014), and the value of spelling practice (Kahn-Horwitz, Schwartz, & Share, 2011) for learning English orthography.

Research supports instruction that activates the three aspects of word identity: semantic, orthographic, and phonological (see Hilte & Reitsma, 2011). In a correlational study of preschool and kindergarten vocabulary practices, Silverman and Crandell (2010) found that teaching the spellings and sounds of vocabulary words was associated with proximal receptive vocabulary learning. In a series of two experiments, Rosenthal and Ehri (2008, 2011) examined the influence of exposing students to the spellings of new vocabulary words they were taught. In the first experiment, which is of most interest to the proposed study, the researchers instructed 20 second graders in a series of learning trials in a laboratory task on two sets of unfamiliar CVC nouns. Researchers defined the words, presented them on word cards with pictures, and embedded them in a series of sentence contexts. For one set of words the students were shown the spellings. On posttests given one day after the learning trials, students learned and remembered pronunciations and meanings of the words that were taught with spelling exposure better than when they simply heard and repeated the words. In a later study with fifth graders, Rosenthal and Ehri (2011) examined the added value of having students pronounce aloud the new vocabulary words they encountered during their silent reading of passages. Compared to students who were asked to simply notice the new underlined vocabulary words (i.e., both groups of students saw the spellings of the words), the students who stopped and pronounced the words aloud had better vocabulary learning; the pronunciation strategy was most beneficial for the lower-skilled readers. In this article we report on a supplemental kindergarten vocabulary intervention designed for English learners to explicitly develop strong word form–meaning connections that support existing core beginning reading curricula.

In summary, potential benefits of the storybook vocabulary approach include a rich story context that supports word learning, in particular when target vocabulary is clearly defined. Drawbacks include less control over the number of word exposures in stories, and isolation of the spoken words from opportunities to decode and spell the words. Potential benefits of the explicit approach include control over multiple exposures to the printed vocabulary words, and practice decoding and spelling the words in list and sentence contexts. Drawbacks include somewhat deprived story and sentence contexts that may limit student engagement and learning.

MODELS OF WORD LEARNING

Extensive research on vocabulary acquisition makes clear that students learn the majority of word meanings through natural reading contexts (Nagy & Anderson, 1984; Nagy & Herman, 1987; Nagy, Herman, & Anderson, 1985). Yet K–1 students are constrained in learning through context given the nature and amount of connected text encountered by students just beginning to learn to read (Baker, Simmons, & Kame'enui, 1998). Many students with limited English lack the sight word vocabulary to take advantage of context to facilitate word recognition in learning new vocabulary (Coady, Carrell, & Nation, 1985; Coady, Magoto, Hubbard, Graney, & Mokhtari, 1993). Students with limited vocabulary and independent reading skills are less likely to close their vocabulary gap solely through independent reading. Research on Matthew effects (Stanovich, 1986) and on the important role of second language vocabulary knowledge in word-reading efficiency and comprehension (Lervag & Aukrust, 2010; Lindsey et al., 2003) supports direct instruction of vocabulary for students with limited English proficiency.

WORD SELECTION PROCEDURES

Fitting vocabulary instruction into the school day requires careful choice of words most worth teaching. This decision depends on both student and word characteristics. High-frequency words are often recommended as a focus for second-language learners (Nation, 2001), and high-frequency root words may bootstrap access and learning of new words from early reading contexts. A second factor in word selection for beginning readers relates to Ehri's (1992) model of the word-reading process. In their reading encounters students build connections in memory between the spellings of words, their pronunciations, and their meanings. In research on adult word learning, exposure to written words improves recognizing words in new contexts, and word learning is improved when spelling is linked to the word meaning (Balass, Nelson, & Perfetti, 2010; Nelson, Balass, & Perfetti, 2005). Word frequency is the characteristic most related to the likelihood a word will be decoded and understood (Adams & Huggins, 1985; Graves, Ryder, Slater, & Calfee, 1987; Juel, 1980). The word corpus for this study was designed to teach well a small set of useful, high-frequency root words that, although familiar to most English-speaking students, may not be known by students with limited English proficiency, and are prerequisites to learning more sophisticated and content words teachers often choose for whole-class vocabulary instruction (Beck, Perfetti, & McKeown, 1982; McKeown, Beck, Omanson, & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985). Words were chosen to be decodable to allow students to apply beginning word-reading and spelling skills. Words were selected from Biemiller's (1999) version of the Dale-Chall (Chall & Dale, 1995) list of 3,000 words commonly known by Grade 4. Biemiller selected 2,300 root words from the list that have the highest utility for instruction. Biemiller and Slonim's (2001) sequence for acquisition of word knowledge supports the usefulness of Living Word Vocabulary (Dale & O'Rourke, 1981) levels for designing a vocabulary scope and sequence. These root words on the Biemiller list were cross-checked with the K–1 Zeno lists (Zeno, Ivens, Millard, & Duvvuri, 1995), which provided a measure (U) of word frequency per million words sampled from texts students are likely to encounter at each grade level. Two criteria were used for word inclusion. First, the word target had to be used frequently and widely at the K–1 grade levels. Second, words had to be decodable based on typical sequence of K–1 alphabetic and phonics skills classroom instruction. This resulted in 184 high-frequency decodable single-syllable words, most of which are nouns or verbs (Table A1, Appendix A includes taught target words and meanings).

A two-cohort randomized trial was used to evaluate the efficacy of explicit vocabulary instruction for EL kindergarteners compared to interactive book reading (IBR) instruction. We anticipated differential effects at posttest, and the primary research questions were:

1. Do students in both conditions exhibit significant gains in vocabulary and beginning phonics skills, both short term (post-treatment Grade K) and longer term (midyear Grade 1)?
2. Do experimental conditions differ on gains?
3. Are treatment effects moderated by initial skill levels (early Grade K), treatment attendance, and treatment fidelity?

METHOD

Initial Study of Connections

In an earlier study Connections was compared to a similar IBR storybook-reading approach (Nelson, Vadasy, & Sanders, 2011). Kindergarten EL students (from Spanish-speaking homes) across six public schools in the Midwest were randomly assigned to one of two small groups within classrooms, which were randomly assigned to Connections or IBR. Small groups comprised two to five students each (depending on the number of EL children at a school), and small-group instruction occurred during the school day, in the classroom, for 30 min per day, four days per week, for the entire school year. Paraeducator tutors were trained and then observed monthly by research staff. Compared to IBR, Connections small groups were estimated to be 4.87 points (or words) higher on posttest reading vocabulary (approximate $d = 1.04$), and 1.78 standard score points higher on posttest decoding ($d = .69$). In a follow-up study six months after intervention ended, results showed that Connections students were still significantly higher than IBR peers on reading vocabulary and decoding (Vadasy, Nelson, & Sanders, 2013).

Participants

Below we describe the present study recruitment, eligibility criteria, random assignment, and attrition during the intervention and follow-up years. Figure A1 in Appendix A illustrates this process for each cohort individually as well as combined.

Recruitment

Sites were public elementary schools located in an urban area of the U.S. Pacific Northwest with large enrollments of English learner (EL) students. Each fall for two years (2011 and 2012), all kindergarten teachers at these schools were given blank consent forms in English and respective home languages to distribute to students for which English was not the primary home language. In Cohort 1, 13 of 14 initially identified sites participated in the study; in Cohort 2, 9 of 12 initially identified sites participated (two of which were new; all sites that participated in Cohort 2 had also participated in Cohort 1), but no student participated in more than one year of intervention. Cohort 1 schools had enrollments averaging 28% receiving EL services (range: 15%–43%), 75% eligible for free/reduced lunch (range: 44%–94%), 84% minority (range: 48%–98%), and 15% receiving special education services (range: 8%–23%). Cohort 2 schools had enrollments averaging 18% receiving EL services, 74% free/reduced lunch eligible, 88% minority, and 16% receiving special education services.

Consents

For Cohort 1, 252 parent consents were received from 38 classrooms across the initial 14 sites; of these, six declined participation and one was received too late. Of the 245 students with consents received prior to testing, eight were dropped prior to randomization due to lack of sufficient EL students within their classroom (a minimum of four students per classroom was required for randomization), 22 were dropped due to lack of a tutor available at one of the schools, and eight moved or were absent for screening. For Cohort 2, 208 parent consents were received for 32 classrooms across 12 sites; of these, four declined participation and five were received too late. Of the 199 students with consents, 21 were dropped prior to randomization due to lack of sufficient EL students within their classroom (resulting in three sites dropped), four moved or were absent for pretesting, and three were dropped due to behavior problems during screening.

Eligibility

We employed a three-pronged approach to eligibility. Within the first month of the academic year, students were initially identified by their classroom teachers as having a home language other than English, even if this was not noted in school records. After receiving parental consent to participate in the study (on which we asked parents to report home language), district scores from the state Washington English Language Proficiency Assessment (WELPA; CTB/McGraw-Hill, 2006) were collected. Any student with consent who scored at Level 1 (Beginning) or 2 (Intermediate) of the four score levels was considered eligible, regardless of other pretests. Second, any student not tested on the WELPA (i.e., if parents refused district testing, or if students registered late) but who scored ≤50th percentile on pretest receptive vocabulary was considered eligible. Our third and final criterion was that classroom teachers were asked to assist in determining whether their EL students would benefit from intervention based on their experience with the student (i.e., with sufficiently low language proficiency). In Cohort 1, of the 207 consented students available for screening, two scored too high on the state test, five scored too high on our vocabulary screener (for students not tested on the state test), and five were removed by their teachers prior to screening as too high in language proficiency; the remaining 195 students were randomly assigned to conditions. In Cohort 2, of the 171 consented students available for screening, five scored too high on the state test, three scored too high on our vocabulary screener, and two were removed by teachers prior to screening as too high in language proficiency; the remaining 161 students were randomly assigned to conditions.

Random Assignment

For both cohorts, students were randomly assigned, within classroom, to small groups of two to three students, with group sizes based on eligible students within each classroom. Small groups were randomly assigned to conditions within classroom, with the added constraint that classrooms with an uneven number of small groups would have an “extra” group randomly assigned to one or the other of the conditions. Cohort 1 included 195 students assigned to 84 small groups, with 42 small groups in each condition (29 dyads and 14 triads in each condition), and Cohort 2 had 161 students assigned to 67 small groups (40 dyads and 27 triads), with 33 small groups in the Connections condition (19 dyads and 14 triads) and 34 small groups in the IBR condition (21 dyads and 13 triads).

Intervention Year (Kindergarten) Attrition

Figure 1A in Appendix A displays attrition for each cohort separately and combined. During the intervention year for Cohort 1, attrition included 15 students (8%): seven moved (four Connections and three IBR), four were removed by their teachers (two per condition), and four (two small groups in the IBR condition) had their tutor withdrawn due to poor fidelity; the tutor's Connections groups were added to another tutor's schedule. Hence, 180 students (91 Connections and 89 IBR) in 81 small groups (42 Connections with an average small group size of 2.17 students, and 39 IBR with an average small group size of 2.28 students) from 31 classrooms across 13 schools was the final sample for Cohort 1. For Cohort 2, attrition included 17 students (11%): eight moved (three Connections and five IBR), three were removed by teachers (two Connections and one IBR), and six were removed due to behavior problems during tutoring (three per condition). Hence, 144 students (72 per condition) in 66 small groups (33 per condition, with an average small group size of 2.18 students) from 24 classrooms across nine schools was the final Cohort 2 sample. To determine whether the final sample was representative of the initial sample, we tested differences between samples on demographic characteristics and pretests using simple t and chi-square tests for each cohort individually and combined. Results showed no significant differences between initial and final samples (ps $> .05$).

Final Kindergarten Sample

The final combined kindergarten sample included 324 EL students in 127 small groups with 163 students in 75 Connections small groups and 161 students in 72 IBR small groups. At pretest, the mean age of students across the sample was 5.61 years ($SD = 0.36$; range $=$ 4.69 to 7.79). Small groups were nested within 27 tutors/tutor teams across 40 classrooms and 13 schools (with some overlap in tutors, classrooms, and schools among the cohorts, but no child participated twice). As shown in Table 1, the combined EL kindergarten sample included 24 languages, with 78% of children scoring in the "intermediate" English language proficiency level on the state test and very few not receiving English language services (6% of parents refused services from the school). Chi-square tests showed no significant differences between cohorts or conditions on any demographic characteristic (all ps $> .05$).

Follow-up Year (First Grade) Attrition

For each cohort, seven months after kindergarten posttest we follow-up tested all students who had participated in the study and were still enrolled in the school district. For Cohort 1, we follow-up tested 162 students, 95% of Connections students ($n = 86$) and 85% of IBR students ($n = 76$). For Cohort 2, we follow-up tested 128 students, 89% of both groups (64 students per condition). Combined, the attrition rate was 10%, with slightly more attrition for the IBR condition (13%) compared to Connections (8%), particularly for Cohort 1. To determine whether the follow-up sample was representative of the intervention year

Table 1. Disaggregated student demographic characteristics

	Connections				IBR				Combined			
	Gr K ($n = 163$)		Gr 1 ($n = 150$)		Gr K ($n = 161$)		Gr 1 ($n = 140$)		Gr K ($n = 324$)		Gr 1 ($n = 290$)	
Characteristic	*N*	*(%)*	*N*	*(%)*	*N*	*(%)*	*N*	*(%)*	*N*	*(%)*	*N*	*(%)*
Female	79	(48%)	71	(47%)	62	(39%)	55	(39%)	141	(44%)	126	(43%)
SPED services	8	(5%)	8	(5%)	6	(4%)	5	(4%)	14	(4%)	13	(4%)
ELL services	152	(93%)	141	(94%)	151	(94%)	131	(94%)	303	(94%)	272	(94%)
State language test												
Beginning	30	(18%)	29	(19%)	28	(17%)	25	(18%)	58	(18%)	54	(19%)
Intermediate	126	(77%)	116	(77%)	126	(78%)	108	(77%)	252	(78%)	224	(77%)
Not tested	7	(4%)	5	(3%)	7	(4%)	7	(5%)	14	(4%)	12	(4%)
Home language												
Spanish	56	(34%)	53	(35%)	50	(31%)	42	(30%)	106	(33%)	95	(33%)
African (7 languages)	57	(35%)	49	(33%)	53	(33%)	46	(33%)	110	(34%)	95	(33%)
Asian (13 languages)	45	(28%)	45	(30%)	48	(30%)	43	(31%)	93	(29%)	88	(30%)
Other (4 languages)	4	(2%)	3	(2%)	5	(3%)	4	(3%)	9	(3%)	7	(2%)
Not reported	1	(1%)	0	(0%)	5	(3%)	5	(4%)	6	(2%)	5	(2%)

sample, particularly given the slight disparity in attrition between conditions, we tested the differences between the samples on demographic characteristics as well as pretests and posttests using simple t and chi-square tests for each cohort individually and combined. Results showed no significant differences between intervention year and follow-up year samples ($ps > .05$). Again, the flowchart in Figure A1 of Appendix A details student recruitment and attrition (the combined cohorts' schools, classrooms, and tutors have some overlap and hence do not directly sum across), and Table 1 provides demographic characteristics of the intervention and follow-up–year samples. Although we label the follow-up year "Grade 1" for brevity (for the remainder of this paper), we note that eight of the 290 students we followed up (3%) had been retained (held back one year) to repeat kindergarten, including three of the 150 in the Connections condition and five of the 140 in the IBR condition. Finally, most of the follow-up sample remained at their original school, with only 11 of the 290 followed up moving to another school within the district, seven of which moved to nonstudy schools ($N = 20$ schools [62 classrooms] at follow-up).

Procedures

Children in both conditions received small-group instruction outside their classroom for 30 min per day, four days per week, for an average of 20 weeks. The primary instructional attributes of each condition are presented in Table A2 in Appendix A (items in bold indicate similarities between the treatment and control conditions) and described below.

The Connections Treatment

Connections materials included a 12- by 17-inch manual used to present lessons to small groups (Nelson & Vadasy, 2007). One new target word was introduced each day (along with two to four related words and cycles of review of previously taught words), a rate of vocabulary instruction at which young students can learn and retain word meanings well (Biemiller, 2003; Clarke, Snowling, Truelove, & Hulme, 2010; Senechal, 1997). For each target word lesson, the student spelled the word orally once, and decoded/pronounced the word eight times. Six activities were used to teach decoding, spelling, and oral production (see Table A3, Appendix A for task details).

The Interactive Book Reading (IBR) Treatment

Students assigned to the IBR condition received instruction in the same target vocabulary provided in the Connections condition. Instruction was provided in the context of reading aloud a storybook in which the target word is featured at least twice. Most of the storybooks used in this study were those selected earlier for use in the pilot study, and were written at the kindergarten/first grade level. The books varied in the number of oral exposures for the target word, and we provided scripted prompts to ensure students interacted with the word an average of three times during the lesson. A list of the storybooks is available from the authors, and the procedures are described in Table A4, Appendix A.

Tutors

All tutors were recruited from their school communities and hired as district employees paid by the schools with funds provided by the research grant. The assignment of tutors to small groups was determined by a combination of school and classroom scheduling, tutor availability, and the number of small groups within classrooms within sites. Each tutor was trained in both treatments and had at least two small groups each. Initially, there were 27 tutors recruited for the study (21 in year 1 but 20 completed; nine of the 20 continued in year 2; and six new in year 2); as noted one tutor in Cohort 1 was dropped from the study and her small groups were reassigned or dropped from the study. In Cohort 1, two tutors worked as a team to cover scheduling for two small groups—one group per treatment condition—but in Cohort 2 were assigned their own small groups (i.e., in the data they are coded as a single tutor team, separate from their individual tutoring in year 2). Thus, the final combined sample included 26 tutors plus one tutor-team, with two tutors working at more than one school (one tutor in each cohort worked at two sites). For both cohorts, tutors averaged four small groups each, with a range of two to six groups with the exception of one tutor (Cohort 2) who had nine small groups at two school sites.

The combined set of 26 tutors (recall two worked as a tutor team in Cohort 1) were mostly female (88%) and white (62%), but varied in age (46% were 45 or older, 35% were 35–44 years old; and the remaining were younger) and educational levels (31% had a bachelor's degree, 77% had some college, and all had a high school diploma or equivalent). Most tutors had previous tutoring experience (69%), including 46% with K–2 experience and 27% with EL tutoring experience. All were required to have state background checks and fingerprinting.

Tutor Training

In both years of the study the researchers conducted a day-long initial training that included an overview of the components for each intervention, modeling how to implement each activity, guided practice in use of the intervention materials, and a review of all intervention materials and recordkeeping procedures. Tutors were introduced to strategies to support and scaffold student responses based on students' developing levels of English proficiency. Researchers provided on-site follow-up training and coaching visits during the first three weeks of the intervention, and as needed. Tutors received feedback linked to the fidelity criteria for each intervention. Strong tutors were videotaped and audiotaped as models for effective instruction and adaptations, and all tutors had access to these tapes. Each week research staff summarized their observations and shared suggestions and strong adaptations with tutors via a weekly e-mail. Tutors were observed on a minimum six additional occasions to assess treatment fidelity. Following each observation, if needed, staff provided corrective feedback. Only one tutor was removed from the project (Cohort 2) after multiple attempts to remediate; half of the tutor's small groups were reassigned to another tutor and the other half were dropped from the study (see previous section regarding attrition).

Treatment Fidelity

Each tutor audiotaped each lesson implemented for each small group. Each month, a 20% random sample of each tutor's audio files was coded for fidelity by a graduate research assistant trained by the third author. A smaller subsample of these randomly selected files (15% random sample, stratified by cohort and month of study) was double-coded by the research assistant and the third author to check coding consistency. Coding indicated the presence of a specific behavior on eight components common to both interventions as well as those unique to each intervention (nine for Connections and five for IBR; a list of all components is given in Appendix A, Table A5). Interrater agreement (percentage of exact matches) showed agreement was extremely high, at 96%, for both common and unique components, and no significant differences in coding agreement between conditions: for common components, Connections agreement M = 94% (SD = 8%) and IBR M = 97% (SD = 7%); for unique components, Connections agreement M = 97% (SD = 7%) and IBR M = 95% (SD = 12%).

Across cohorts and groups, there were 844 coded fidelity recordings (Connections average recordings per small group M = 5.73, SD = 2.46; IBR M = 5.75, SD = 2.64). Across recordings, the observed common component fidelity mean was 95% (SD = 11%), and unique component fidelity mean was 95% (SD = 10%). Multilevel models with observations nested within small groups within tutors showed no significant difference between conditions on unique components fidelity (Connections M = 96%, SD = 9%; IBR M = 95%, SD = 12%); however, there was a difference of 2% (approximately 0.25 standard deviations) between conditions on common components fidelity (Connections M = 96%, SD = 8%; IBR M = 94%, SD = 11%), with Connections fidelity higher than IBR ($p < 0.01$). Finally, it is worth noting that substantial variation between tutors for both unique and common component fidelity was observed in our fidelity analyses: tutor intraclass correlations (ICCs) were 0.22 and 0.13 for unique and common components, respectively. Small groups, however, did not vary as much within tutors, with ICCs <0.01 and 0.06 for unique and common components, respectively. Given that tutors appeared to have more difficulty implementing common components for IBR, we saved model-implied small-group common component fidelity for our treatment effects models.

Treatment Attendance

Throughout the intervention, tutors recorded attendance for each child in each small group (one target word lesson per day of tutoring). The observed student mean was 73.10 days of tutoring out of 80 possible (SD = 7.28, range = 40–80 days; Connections M = 72.82, SD = 7.63; IBR M = 73.39, SD = 6.92). Multilevel models of students nested within small groups within tutors showed no significant differences between cohorts or conditions (at both the student and small-group levels); however, a very large proportion of between-tutor variance was observed on attendance, with an estimated tutor ICC = 0.66 (whereas small groups did not vary much within tutors on attendance: ICC < 0.01). Given the large variability in attendance, we incorporate these data in our forthcoming treatment effects models.

Student Assessments

For each cohort, students were pretested in the fall prior to treatment and posttested in the spring just after treatment, approximately seven months apart, during their kindergarten year, and were followed up approximately seven months after posttest, in the winter of the following year. For each test wave, students were tested individually in English by trained testers unaware of group assignment. To ensure comparability with previous research as well as to determine gains not related to normative developmental changes, age-based standard scores (normative population M = 100, SD = 15) were used for all analyses with the exception of one experimenter-developed measure. All measures were administered using identical test forms at each test wave. Descriptions of assessments, including sample reliabilities (Cronbach's alpha), are as follows:

1. *Receptive vocabulary* was measured using the Peabody Picture Vocabulary Test-IIIA (PPVT-IIIA; Dunn & Dunn, 2006). For this test, children select a picture that best illustrates the meaning of an orally presented stimulus word and testing is discontinued after the student misses 8 out of 12 items within a grade level set. The maximum number of words is 204. Test manual split-half reliability is 0.76 for five-year-olds and .80 for six-year-olds. Across cohorts, median sample reliabilities were 0.94, 0.92, and 0.95 at pretest, posttest, and follow-up, respectively.
2. *Reading vocabulary* was assessed using an experimenter-developed 25-item curriculum-based measure (CBM) of target word reading vocabulary. For each item, the tester showed the student a group of three words, one of which was the target. The tester read aloud a definition, for example, "Point to the word that means *to move your head up and down*," and the student was asked to read the three word items (e.g., *hop*, *nod*, *wink*) and point to the word matching the definition. Practice items were repeated until the student understood the directions, and a prompt was given after 3 sec (with up to 10 sec for student to respond). All items are administered and the score is the percent of items correct. Median sample reliabilities were 0.57, 0.86, and 0.80 at pretest, posttest, and follow-up, respectively. To establish validity evidence, we computed simple, unadjusted correlations between the CBM measure and receptive vocabulary and decoding measures. Zero-order correlations with receptive vocabulary were found to be $r = 0.19$, 0.49, and 0.55 at pretest, posttest, and follow-up, respectively (ps $< .001$); its correlations with decoding were $r = 0.06$, 0.64, and 0.60 at pretest, posttest, and follow-up, respectively (ps $< .001$, except pretest $p > .05$). Moreover, partial correlations were computed to establish the relationship between the CBM and receptive vocabulary *after controlling for decoding*, which were $r = 0.18$, 0.30, and 0.43 at pretest, posttest, and follow-up, respectively (ps $< .001$), indicating a positive relationship between the two vocabulary measures beyond the CBM relationship with decoding.
3. *Decoding* was measured using the Woodcock Reading Mastery Test-Revised/Norm Referenced (WRMT-R/NU; Woodcock, 1987/1998) Word Attack subtest. The test includes 45 nonwords that follow typical English orthographic rules and that steadily increase in difficulty. Students are asked to read each nonword individually, with testing discontinued after six consecutive incorrect responses. Test manual split-half reliability for kindergartners is 0.94. Median sample reliabilities were 0.89, 0.92, and 0.95 at pretest, posttest, and follow-up, respectively.
4. *Spelling* was measured at pretest and posttest using the Wide Range Achievement Test-4 (WRAT-4; Wilkinson & Robertson, 2006) Spelling subtest, which requires name writing and 13 dictated letters, and spelling of 42 dictated, increasingly difficult words. Testing is discontinued after 10 consecutive errors. Test manual internal consistency reliability coefficient for kindergarteners is .94. Median sample reliabilities were 0.84, 0.82, and 0.84 at pretest, posttest, and follow-up, respectively.

Analysis Plan

A multilevel hierarchical modeling approach was adopted for testing the research questions. All models were estimated with maximum likelihood in *HLM7*. Except for the curriculum-based measure in which percentage correct was used, all analyses were conducted on norm-referenced standard scores. Below we detail our preliminary and final analysis models.

Nesting Structure

We evaluated intraclass correlations (ICCs) on outcomes and gains, as well as testing for cohort or group differences prior to treatment and checking the general growth patterns. First, we determined the optimal nesting structure for data analyses by comparing ICCs computed from variance component estimates of three-level intercept-only models in which students ($N = 324$) were nested within small groups ($N = 147$), within either (a) schools ($N = 13$) or (b) tutors ($N = 27$ tutors/tutor-teams). For brevity, we ignored classroom membership ($N = 40$ classrooms during intervention and $N = 62$ during follow-up) as an intermediate level because we had no research questions pertaining to classroom-level predictors, and classroom variance would be absorbed into school- or tutor-level variance. Comparison of the sum of each model's ICCs showed near equivalence for each measurement occasion as well as on gain scores. Because we found that schools accounted for less variance in outcomes compared with tutor ICCs, and because we previously found heterogeneity among tutors on fidelity, we concluded that tutors would be the optimal highest structure level for these data.

Pretest Cohort Differences

Next, we evaluated whether the two cohorts differed on pretests (as already mentioned, there were no significant demographic differences) using 3-level models in which the cohort was tested at the small-group level (results did not differ when tested at the individual level). We found one difference between cohorts on pretest receptive vocabulary: the second cohort was an estimated 4.90 points lower than the first cohort at pretest (slightly less than a third of a standard deviation difference), and this difference was similar at posttest. Hence, we controlled for small groups' cohort membership in all final analyses. One reason for the difference may be that immigrant populations such as ours change annually, which could reflect subtle changes in socioeconomic backgrounds affecting English vocabulary knowledge. Between cohort 1 and 2, children from Spanish-speaking homes increased from 32% to 34% as did children from homes speaking Chinese and SE Asian languages (from 27% to 31%), whereas children from homes speaking an African language decreased from 37% to 31%.

Pretest Condition Differences

We also examined whether the two treatments differed on any pretest (there were no demographic differences). Again we conducted three-level models in which students were nested within small groups, within tutors (collapsed across cohorts), with treatment tested at the small-group level. Results indicated that the Connections condition was slightly higher than the IBR condition on pretest decoding and spelling (0.25 and 0.26 standard deviations, respectively); no other significant differences were observed. Unfortunately, randomization does not guarantee that experimental conditions 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 non-equivalence to occur by chance in a multi-cohort, cluster-randomized design such as ours is greater because we are dealing with a relatively smaller number of units to be randomized (small groups, not students), and further, our units for randomization were even smaller due to the two-cohort nature of the study. These nonzero differences are not detectable in separate cohort analyses; however, with the combined cohort data we were able to detect even very small differences of 0.25 standard deviations. To minimize the selection bias introduced by nonequivalent conditions, our final analysis plan focuses on gains rather than absolute posttest levels, and additionally, we use student pretests as a covariate to control for any effects of initial treatment differences on gains.

Final Models

Our final models were three-level models in which student gains were nested within small groups, nested within tutors/tutor-teams. All treatment effects were tested at the small-group level since that was the unit of randomization. For each gain outcome, we conducted a series of three models in a fashion similar to sequential regression. Model 1 estimated gains, adjusted for small-group and tutor membership, to determine whether growth across both experimental conditions was significantly different from zero (intercept-only models). Model 2 estimated treatment effects on gains (at the small-group level) adjusted for cohort (at the small-group level), respective pretest (at the student level), attendance (at the student level), and common treatment component fidelity (at the small-group level). Model 3 added two-way interactions between treatment and the pretest, attendance, and fidelity covariates. For ease of interpretation, cohort and experimental conditions were effect coded (cohort: $+1$ = last, -1 = first cohort, and condition: $+1$ = Connections and -1 = IBR), and each of the other covariates were standardized in z-scores. Approximate effect sizes (denoted d^*) were computed for cohort and treatment differences as twice the coefficient estimate (twice due to effect coding of conditions) divided by the square root of the sum of the variance component estimates (i.e., approximate standard deviation); all other predictors' effect sizes were computed as the coefficient estimate divided by the approximate standard deviation. In other words, the effect size for cohort or condition may be interpreted as the approximate distance in gains between the cohorts or conditions, in standard deviations; the effect sizes for other covariates may be interpreted as the change in gains for every standard deviation increase in the predictor.

Adjusted Alpha Level for Analyses on Multiple Outcomes. With four outcomes analyzed (receptive language, reading vocabulary, decoding, and spelling) for each of the two sets of gains, we run the risk of inflating Type I error beyond the nominal 0.05 level. Because each of the two sets of gains answers two distinct research questions, we treat each set of gains as its own *family*, and control for Type I error inflation *family-wise* using the Dunn-Sidak (DS) procedure. Specifically, the DS procedure adjusts the per-outcome alpha level by algebraically rearranging the total Type I error rate formula, which is: $1 - (1 - \text{per-outcome } \alpha)^{\mathrm{m}}$, where m = number of outcomes, to an adjusted level equal to: $1 - (1 - \text{total desired } \alpha)^{1/\mathrm{m}}$. Given that we had m = 4 outcomes, our adjusted alpha level for a given set of gain models is 0.0127, rather than .05. Only results that are statistically significant at the adjusted level will be interpreted for discussion (and boldfaced in our results tables); however, because readers may wish to examine certain outcome results in isolation of other outcomes, in our results tables we have reported both unadjusted (using asterisks) as well as adjusted (using boldface for all tests significant at the 0.0127 level) test results. We note that Models 1 and 2 are subsumed in Model 3 and so Model 3 is the ultimate focus of our results section; additionally, the number of independent variables (predictors) tested within a model does not relate to Type I error inflation, because these tests of coefficients are protected by the overall model-based degrees of freedom and corresponding variance estimates. Finally, we also note that the What Works Clearinghouse's (2014) *Procedures and Standards Handbook Version 3.0* recommends use of the Benjamini-Hochberg (BH) procedure for controlling Type I error, which is a sequential method to control the "false discovery rate." Because readers would likely have a difficult time following this logic in our results, particularly given the multiple predictors in each of our models that would have to be compared against respective predictors in respective models only, we opted to use the simpler DS procedure, which can be slightly conservative; nevertheless, we note that our substantive findings would not change had we used the BH procedure, with only one minor exception in our forthcoming immediate effects models that is explicitly noted (about pretest effects).

RESULTS

Descriptive Statistics

Disaggregated student assessment means and standard deviations for pretests, posttests, and follow-up tests for each condition are provided in Table 2; for reader interest zero-order correlations for each condition are provided in Table B1, Appendix B. At pretest, based on norm-referenced scores, the sample averaged in the 10th percentile in receptive vocabulary, 35th percentile in decoding, and 27th percentile in spelling. At posttest, the sample averaged in the 17th, 62nd, and 46th percentiles, respectively, exhibiting substantial gains from pretest. At follow-up, the sample averaged in the 18th, 65th, and 49th percentiles, respectively (i.e., maintaining vocabulary and increasing slightly in decoding and spelling).

Pretest Differences Between Cohorts and Conditions

As already noted in the Analysis Plan, preliminary analyses revealed that cohort 2 was significantly lower than cohort 1 on pretest receptive vocabulary (approximately one third of a standard deviation), and that the Connections group was significantly higher than IBR on pretest decoding and spelling (approximately one fourth of a standard deviation). As such, our forthcoming treatment effects models adjust for cohort as well as pretest, and further, focus on gains as outcomes, rather than posttest levels.

Intraclass Correlations

Table 3 displays intraclass correlations (ICCs) computed using variance components estimates from unconditional three-level models for immediate (pretest to posttest) and longer-term (pretest to follow-up) gains. Results show the median small-group ICC was 0.13 for immediate gains and 0.01 for longer-term gains, and the median tutor ICC was .02 and .01, respectively. Taken together, small groups and tutors accounted for an average of 14% of the variance in immediate student gains and 3% of the variance in longer-term student gains.

Table 2. Disaggregated student assessment descriptive statistics

	Connections						IBR					
	Gr K ($n = 163$)				Gr 1 ($n = 150$)		Gr K ($n = 161$)				Gr 1 ($n = 140$)	
	Pretest		Posttest		Follow-up		Pretest		Posttest		Follow-up	
Measure	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Receptive vocabulary	72.13	(16.56)	81.21	(14.67)	83.04	(13.59)	71.34	(16.72)	80.32	(13.49)	81.55	(12.13)
Reading vocabulary %	0.33	(0.10)	0.62	(0.23)	0.63	(0.20)	0.34	(0.09)	0.49	(0.19)	0.58	(0.18)
Decoding	95.20	(4.71)	108.71	(11.56)	109.35	(12.96)	94.32	(2.32)	103.07	(11.92)	105.22	(14.08)
Spelling	88.95	(13.86)	100.25	(13.90)	102.05	(15.03)	85.36	(13.27)	95.45	(12.96)	97.66	(15.08)

Note. Receptive vocabulary = standard score of Peabody Picture Vocabulary Test-IIIA; Reading vocabulary = percent correct out of 25 items on curriculum-based measure of root word vocabulary; Decoding = standard score of Woodcock Reading Mastery Test-Revised/Normative Update Word Attack subtest; Spelling = standard score of Wide Range Achievement Test-4 Spelling subtest.

Table 3. Estimated intraclass correlations

	Outcome			
Source	Receptive Vocabulary	Reading Vocabulary	Decoding	Spelling
Immediate gains				
Tutors	0.11	0.03	<.01	<.01
Small groups	0.01	0.18	0.08	0.24
Longer-term gains				
Tutors	0.02	0.01	0.01	0.01
Small groups	<.01	<.01	0.02	0.06

Note. Maximum likelihood estimates computed from variance component estimates in 3-level unconditional models. Immediate gains = kindergarten fall pretest to kindergarten spring posttest (N = 324 students within 147 small groups and 27 tutors); Longer-term gains = kindergarten fall pretest to first-grade winter follow-up test (N = 290 students within 145 small groups and 27 tutors); Receptive vocabulary = standard score of Peabody Picture Vocabulary Test-IIIA; Reading vocabulary = percent correct out of 25 items on curriculum-based measure of root word vocabulary; Decoding = standard score of Woodcock Reading Mastery Test-Revised/Normative Update Word Attack subtest; Spelling = standard score of Wide Range Achievement Test-4 Spelling subtest.

Immediate (Pretest–Posttest) Gains

Model 1: Gains Only

Results from our immediate gains models are provided in Tables 4 and 5 for vocabulary and literacy outcomes, respectively. Across outcomes, Model 1 results (unconditional models) show that the students across both groups made significant gains from pretest to posttest. For the three norm-referenced measures (receptive vocabulary, decoding, and spelling), mean gains averaged approximately 10 points, which translates to two thirds of the population standard deviation for these measures. On the curriculum-based reading vocabulary measure, the increase was 22% (from pretest averaging 33%); in other words, by posttest, students averaged just over half of the 25 root word items correct on reading vocabulary.

Model 2: Treatment and Covariate Effects

Results from Model 2 (Tables 4 and 5), which included cohort, condition, pretest, attendance, and fidelity, showed several key findings. Most important, we found significant treatment condition differences, adjusted for all the other covariates, favoring the Connections condition over IBR on pretest–posttest gains in reading vocabulary and decoding, with medium-sized effects of $d^* = 0.64$ (12.6% difference in gains) and 0.45 (4.92-point difference in gains), respectively. No significant differences were detected for receptive vocabulary or spelling. Model 2 results also indicated several other findings. First, cohort negatively predicted two gain outcomes: specifically, cohort 2 had lower gains than cohort 1 on reading vocabulary and decoding (.46 and .42 standard deviations, respectively), which may indicate that lower receptive vocabulary skills relates to lower growth in these areas (since cohort 2 had lower initial receptive vocabulary than cohort 1). Second, and as may be expected, results showed that, regardless of treatment condition, pretest negatively predicted gains for all outcomes except decoding (based on the Dunn-Sidak adjusted alpha level; if

Table 4. Model results for immediate gains on vocabulary outcomes

	Receptive Vocabulary					Reading Vocabulary (CBM)				
	Model 1	Model 2		Model 3		Model 1	Model 2		Model 3	
Fixed Effects	Coeff	Coeff	d*	Coeff	d*	Coeff	Coeff	d*	Coeff	d*
Immediate mean gain	**8.86*****	**8.86*****		**8.80*****		**0.217*****	**0.214*****		**0.217*****	
Cohort (1 = last)		−0.91	−.20	−1.00	−.22		**−0.045*****	−.46	**−0.044*****	−.46
Condition (1 = CONN)		0.15	.03	0.13	.03		**0.063*****	.64	**0.063*****	.65
Student pretest		**−6.05*****	−.66	**−6.00*****	−.66		**−0.069*****	−.35	**−0.072*****	−.37
Student attendance		**1.74*****	.19	**1.71****	.19		**0.058*****	.30	**0.054*****	.28
Small-group fidelity		0.67	.07	0.72	.08		**0.029****	.15	**0.029****	.15
Condition * pretest				0.87					0.016	
Condition * attend				−0.55					**0.026***	
Condition * fidelity				0.03					−0.005	
Random Effects	Var	Var		Var		Var	Var		Var	
Tutors	**12.23*****	**9.64*****		**10.21*****		0.002	0.001		0.001	
Small groups	1.31	0.26		0.00		**0.009****	0.002		0.002	
Residual	109.49	74.72		73.72		0.042	0.036		0.035	

Note. Immediate gain = kindergarten fall pretest to kindergarten spring posttest (N = 324 students within 147 small groups and 27 tutors). Receptive vocabulary = standard score of Peabody Picture Vocabulary Test-IIIA; Reading vocabulary = percent correct out of 25 items on curriculum-based measure of root word vocabulary; Cohort = effect coded (+1 = last, −1 = first cohort); Condition = effect coded (+1 = Connections, −1 = IBR treatment); Pretest = student fall score on corresponding outcome, standardized into z-scores; Attendance = number of tutoring sessions students attended, standardized into z-scores; Fidelity = small-group mean fidelity across the intervention period, standardized into z-scores. *unadjusted $p < .05$, **unadjusted $p < .01$, **unadjusted $p < .001$; boldface = statistically significant (Dunn-Sidak procedure) after adjusting for four outcomes analyzed for each gain set (significance criterion: $p < .0127$). (Dunn-Sidak adjustment procedure yields same substantive results as Benjamini-Hochberg adjustment procedure.)

Table 5. Model results for immediate gains on early literacy outcomes

	Decoding					Spelling				
	Model 1	Model 2		Model 3		Model 1	Model 2		Model 3	
Fixed Effects	Coeff	Coeff	d*	Coeff	d*	Coeff	Coeff	d*	Coeff	d*
Immediate mean gain	**11.13*****	**10.89*****		**10.87*****		**10.61*****	**10.50*****		**10.35*****	
Cohort (1 = last)		**−2.27*****	−.42	**−2.24*****	−.41		−0.82	−.17	−0.89	−.19
Condition (1 = CONN)		**2.46*****	.45	**2.49*****	.46		0.89	.19	0.89	.19
Student pretest		−1.02*[a]	−.09	**−1.34*****	−.12		**−4.79*****	−.50	**−4.74*****	−.50
Student attendance		**2.27****	.21	**2.24****	.21		0.84	.09	0.88	.09
Small-group fidelity		1.02	.09	1.04	.09		**1.80*****	.19	**1.91*****	.20
Condition * pretest				0.52					0.28	
Condition * attend				0.16					−0.60	
Condition * fidelity				−0.21					0.41	
Random Effects	Var	Var		Var		Var	Var		Var	
Tutors	0.05	0.01		0.02		0.01	0.31		0.98	
Small groups	11.20	2.67		2.76		**27.29*****	**28.19*****		**26.83*****	
Residual	122.06	116.50		116.16		86.92	62.53		62.61	

Note. Immediate gain = kindergarten fall pretest to kindergarten spring posttest ($N = 324$ students within 147 small groups and 27 tutors). Decoding = standard score of Woodcock Reading Mastery Test-Revised/Normative Update Word Attack subtest; Spelling = standard score of Wide Range Achievement Test-4 Spelling subtest; Cohort = effect coded (+1 = last, −1 = first cohort); Condition = effect coded (+1 = Connections, −1 = IBR treatment); Pretest = student fall score on corresponding outcome, standardized into z-scores; Attendance = number of tutoring sessions students attended, standardized into z-scores; Fidelity = small-group mean fidelity across the intervention period, standardized into z-scores. *unadjusted $p < .05$, **unadjusted $p < .01$, **unadjusted $p < .001$; boldface = statistically significant (Dunn-Sidak procedure) after adjusting for four outcomes analyzed for each gain set (significance criterion: $p < .0127$); [a] = would be statistically significant according to the Benjamini-Hochberg per-comparison adjustment, after sorting all pretest predictor p-values across all four immediate gains Model 2 (pretest effect observed p-value = .023 for decoding, the largest of all four pretest p-values, which is less than $\alpha/1 = .05$). (Otherwise, note that Dunn-Sidak adjustment procedure yields same substantive results as Benjamini-Hochberg adjustment procedure.)

no adjustment is made or if Benjamini-Hochberg adjustment is used, results would show that pretest decoding negatively predicts decoding gains as well), indicating that students who had relatively high pretests had lower gains. Third, regardless of treatment condition, student attendance was positively predictive of both types of vocabulary gains as well as decoding gains: for every standard deviation increase in student attendance, there was a predicted increase in gains of 1.74 points in receptive vocabulary ($d^* = 0.19$), 6% in reading vocabulary ($d^* = 0.30$), and 2.27 points in decoding ($d^* = 0.21$). Finally, small-group treatment fidelity was found to be predictive of reading vocabulary and spelling gains: for every standard deviation increase in fidelity, a 3% increase in reading vocabulary gains ($d^* = 0.15$) and a 1.80-point increase in spelling gains was predicted ($d^* = 0.19$).

Model 3: Treatment Interactions

In our final model for immediate gains (Model 3, Tables 4 and 5), results showed the same substantive findings for covariate main effects observed for Model 2 with the exception of pretest decoding's effect on decoding gains (which became statistically significant in Model 3). Moreover, only one small interaction was detected across all four outcomes: specifically, there was an interaction between condition and attendance on reading vocabulary gains (Table 4). To understand the nature of this interaction, we computed model-implied values of gains for relatively low (−1 standard deviation) and relatively high (+1 standard deviation) levels of attendance by experimental condition, and found that the advantage for the Connections condition to have greater gains than IBR was bolstered as attendance increased. Specifically, there was a 7% advantage on gains for relatively lower attendance (a level of approximately 66 days), 13% advantage for average attendance (approximately 73 days), and an 18% advantage for higher attendance (maximum attendance of 80 days).

Longer-Term (Pretest–Follow-up) Gains

Model 1: Gains Only

The same models described above were used on longer-term gains (from pretest in early Grade K to follow-up test in Grade 1). As shown for Model 1 (Tables 6 and 7), longer-term gains were significantly greater than zero across all vocabulary and early literacy outcomes. Inspection of the observed means (Table 2) illustrates a lack of any decline from kindergarten posttest (end of year) to midyear of first grade. Indeed, post-hoc tests of gains (using three-level unconditional models) from posttest to follow-up showed significant growth, albeit small, on all outcomes (an estimated gain of 1.43 points on receptive vocabulary, 5% on reading vocabulary, 1.13 points on decoding, and 1.77 points on spelling).

Model 2: Treatment and Covariate Effects

Importantly, the previous treatment differences on longer-term gains were consistent with the immediate gains models (see Model 2 across Tables 6 and 7), again showing an advantage for the Connections condition over the IBR condition on reading vocabulary and decoding. This said, the longer-term gains showed effect sizes to be much smaller, with $d^* = 0.29$ and 0.27 for reading vocabulary and decoding, respectively. Again, we conducted post-hoc models of gains between posttest kindergarten and follow-up Grade 1. Results showed that the IBR group made significantly higher gains than Connections did

Table 6. Model results for longer-term gains on vocabulary outcomes

	Receptive Vocabulary					Reading Vocabulary (CBM)				
	Model 1	Model 2		Model 3		Model 1	Model 2		Model 3	
Fixed Effects	Coeff	Coeff	d*	Coeff	d*	Coeff	Coeff	d*	Coeff	d*
Longer-term mean gain	**10.57*****	**10.62*****		**10.38*****		**0.271*****	**0.266*****		**0.265*****	
Cohort (1 = last)		0.31	.06	0.08	.02		**−0.035****	−.38	**−0.034****	−.38
Condition (1 = CONN)		0.34	.07	0.34	.07		**0.026****	.29	**0.026****	.29
Student pretest		**−8.31*****	−.86	**−8.34*****	−.87		**−0.081*****	−.44	**−0.083*****	−.46
Student attendance		**1.44***	.15	1.32*	.14		**0.035****	.19	**0.032***	.17
Small-group fidelity		0.08	.01	0.28	.03		0.017	.09	0.020*	.11
Condition * pretest				**1.45****					0.014	
Condition * attend				−0.36					0.018	
Condition * fidelity				0.76					0.013	
Random Effects	Var	Var		Var		Var	Var		Var	
Tutors	2.60	0.02		0.19		0.001	<.001		<.001	
Small groups	0.32	0.16		0.11		<.001	<.001		<.001	
Residual	162.29	93.82		91.23		0.042	0.034		0.033	

Note. Longer-term gain = kindergarten fall pretest to first-grade winter follow-up test (N = 290 students within 145 small groups and 27 tutors). Receptive vocabulary = standard score of Peabody Picture Vocabulary Test-IIIA; Reading vocabulary = percent correct out of 25 items on curriculum-based measure of root word vocabulary; Cohort = effect coded (+1 = last, −1 = first cohort); Condition = effect coded (+1 = Connections, −1 = IBR treatment); Pretest = student fall score on corresponding outcome, standardized into z-scores; Attendance = number of tutoring sessions students attended, standardized into z-scores; Fidelity = small-group mean fidelity across the intervention period, standardized into z-scores. *unadjusted $p < .05$, **unadjusted $p < .01$, **unadjusted $p < .001$; boldface = statistically significant (Dunn-Sidak procedure) after adjusting for four outcomes analyzed for each gain set (significance criterion: $p < .0127$). (Note that Dunn-Sidak adjustment procedure yields same substantive results as Benjamini-Hochberg adjustment procedure.)

Table 7. Model results for longer-term gains on early literacy outcomes

	Decoding					Spelling				
	Model 1	Model 2		Model 3		Model 1	Model 2		Model 3	
Fixed Effects	Coeff	Coeff	d*	Coeff	d*	Coeff	Coeff	d*	Coeff	d*
Longer-term mean gain	**12.42*****	**12.41*****		**12.30*****		**12.63*****	**12.53*****		**12.50*****	
Cohort (1 = last)		−0.32	−.05	−0.28	−.04		−0.43	−.07	−0.39	−.07
Condition (1 = CONN)		**1.76****	.27	**1.81****	.28		0.57	.10	0.57	.10
Student pretest		**−1.59****	−.12	**−2.22*****	−.17		**−4.85*****	−.41	**−4.84*****	−.41
Student attendance		**2.04****	.16	**1.98****	.15		0.55	.05	0.49	.04
Small-group fidelity		**1.60****	.12	**1.71****	.13		1.78*	.15	1.76*	.15
Condition * pretest				**1.18***					0.51	
Condition * attend				0.18					−0.17	
Condition * fidelity				−0.07					−0.34	
Random Effects	Var	Var		Var		Var	Var		Var	
Tutors	1.52	0.04		0.05		1.25	1.89		1.95	
Small groups	3.86	1.35		0.46		9.54	11.94		11.06	
Residual	174.16	167.13		167.00		155.58	129.19		129.58	

Note. Longer-term gain = kindergarten fall pretest to first-grade winter follow-up test (N = 290 students within 145 small groups and 27 tutors). Decoding = standard score of Woodcock Reading Mastery Test-Revised/Normative Update Word Attack subtest; Spelling = standard score of Wide Range Achievement Test-4 Spelling subtest; Cohort = effect coded (+1 = last, −1 = first cohort); Condition = effect coded (+1 = Connections, −1 = IBR treatment); Pretest = student fall score on corresponding outcome, standardized into z-scores; Attendance = number of tutoring sessions students attended, standardized into z-scores; Fidelity = small-group mean fidelity across the intervention period, standardized into z-scores. *unadjusted $p < .05$, **unadjusted $p < .01$, **unadjusted $p < .001$; boldface = statistically significant (Dunn-Sidak procedure) after adjusting for four outcomes analyzed for each gain set (significance criterion: $p < .0127$). (Note that Dunn-Sidak adjustment procedure yields same substantive results as Benjamini-Hochberg adjustment procedure.)

during Grade 1 on reading vocabulary and exhibited the same trend for decoding, in effect "catching up" to the Connections group (no differences in gains for receptive vocabulary or spelling). Specifically, the predicted mean gain on reading vocabulary during this growth period was 9% for IBR whereas the Connections condition students gained less than 1% on reading vocabulary ($p < .001$). The predicted mean gain on decoding was 2.33 points for IBR and 0.04 points for Connections; however, this difference was not significant ($p = .056$).

Other results from this model for longer-term gains (Model 2, Tables 6 and 7), compared with shorter-term gains (Tables 4 and 5), showed that pretest persisted in its negative association with gains (across all four outcomes), as did attendance with receptive vocabulary, reading vocabulary, and decoding gains (regardless of treatment condition). However, cohort and small group fidelity were not consistently predictive of growth from immediate to longer-term gains: cohort was no longer predictive of gains in decoding (only predictive of reading vocabulary) and fidelity was no longer significantly predictive of gains in reading vocabulary or spelling but became predictive of longer-term gains in decoding.

Model 3: Treatment Interactions

The results observed in Model 2 (on longer-term gains) did not exhibit much change when interaction effects were added to our final model, Model 3. Further, the only interactions detected were among condition and pretest on longer-term gains in receptive vocabulary and decoding. When we computed the model-predicted values for each of these outcomes, we found that the interaction on decoding gains was ordinal (increasing positive treatment effects for Connections over IBR for students with higher pretest levels), but that the interaction found for vocabulary gains was disordinal (fully crossed). Specifically, students who were relatively low in receptive language (-1 standard deviation, which is $<$1st percentile in our sample) made greater gains in the IBR condition compared with the Connections condition (2.22 points more on gains, $d^* = 0.23$), whereas students with relatively higher pretest receptive language levels ($+1$ standard deviation, which is approximately the 21st percentile) made greater gains in the Connections condition (3.57 points more on gains, $d^* = 0.37$). However, recall that there was no main effect detected on receptive vocabulary gains for either immediate or longer-term gains. For all practical purposes, this finding simply indicates that EL students who are extremely low in receptive vocabulary may benefit in language development more from IBR treatment, and students who are higher than the typical EL student may benefit more from Connections. Students in our sample who scored one standard deviation or more below average on pretest receptive vocabulary ($n = 59$) had a mean of 14 raw score points (range: 0–26 points) and only a small handful ($n = 4$) had a score of zero points. Comparatively, those who were at least one standard deviation above average on pretest ($n = 51$) had a mean of 67 raw score points (range: 52–84 points).

DISCUSSION

In this two-cohort cluster-randomized study we compared two approaches to supplementing vocabulary instruction for kindergarten English learners: an explicit approach to build semantic, orthographic, and phonological connections for English vocabulary learning, and an interactive storybook reading approach. At pretest, children averaged at the 10th percentile in receptive vocabulary. Children in both interventions made progress learning taught vocabulary and on norm-referenced receptive vocabulary, decoding, and spelling measures, although it is difficult to interpret these gains without a no-treatment control comparison. Children receiving Connections made significantly greater gains in taught reading vocabulary and in decoding. Similar to findings reported by others, children entering with higher vocabulary acquired vocabulary more easily than children entering with lower levels (Ewers & Brownson, 1999; Robbins & Ehri, 1994; Senechal, Thomas, & Monker, 1995). In the present sample, pretests negatively predicted gains, and EL children's initial skills had an independent effect on selected outcomes. Decoding posttest gains were higher for children in Connections with lower decoding pretests, and longer-term decoding gains were higher for children in Connections with higher decoding pretest. The mean scores for decoding and spelling at posttest and at follow-up were at or close to grade level for both groups, similar to findings on the development of word-reading skills in EL students reported by others (e.g., Lesaux, Koda, Siegel, & Shanahan, 2006). We found that student attendance predicted all gains; fidelity of instruction predicted posttest and follow-up spelling and follow-up decoding. The patterns of findings in this study (significantly greater posttest gains for Connections in reading vocabulary and decoding, and sustained advantage for these measures at follow-up) were similar to those in the pilot study.

Interestingly, EL children in the storybook reading condition exhibited better longer-term receptive vocabulary gains than those in the Connections condition if their incoming receptive vocabulary skills were quite low (i.e., <1st percentile on national norms for the PPVT-IIIA). However, EL children with somewhat better incoming skills (i.e., > 20th percentile) demonstrated greater gains in the longer term if they received the Connections treatment. Storybook reading approaches such as IBR may provide more varied and rich context for word learning, as well as discourse opportunities that benefit language growth for English learners who enter kindergarten with lower English language skills. We note the strength of the IBR condition in this study that included features earlier found to support vocabulary acquisition in story-reading contexts, including explicit explanations of words, active questioning and rephrasing, word practice and review, and multiple word exposures (Marulis & Neuman, 2010). Children who entered kindergarten with higher English language skills in the Connections treatment may have the language skills to fully benefit from the intervention activities and language interactions with an English-speaking adult, and better transfer learning of taught vocabulary to more general vocabulary. Findings suggest the Connections approach benefits most the EL kindergarteners who entered with limited alphabet and decoding skills. The practice and support in blending words offered significantly greater immediate and longer-term gains in decoding. IBR instruction appears best suited for the children who entered kindergarten with lower receptive vocabulary for whom the intervention had longer-term general vocabulary advantages. The lack of differential instruction effects on spelling may be due to the very brief single oral spelling of the target word in each Connections lesson. Added spelling opportunities, including written spelling of taught vocabulary, may afford children practice writing letters to build stronger orthographic-semantic connections, and this is a feature we will examine in future research.

Children averaged at the 17th percentile in receptive vocabulary at posttest. As others have reported, although neither intervention closed the gap in English vocabulary knowledge for EL kindergartners (Marulis & Neuman, 2010), the proximal learning benefits for Connections suggests small details of instruction that may enhance vocabulary learning. These include practice decoding and reading the target words, oral and written spelling of words, and pronouncing the words. An appealing feature of the decoding and spelling practice is its brevity: it required little effort or special design. A strength of the study is its comparative design featuring two viable and developmentally appropriate methods for teaching vocabulary. Findings for IBR add to an extensive research base on children's vocabulary learning from listening to stories along with teacher explanations of target words. This approach may afford the concrete, diverse, and complex contexts that support the longer-term language growth we found for IBR children entering with lower receptive vocabulary (see Aukrust & Rydland, 2011).

LIMITATIONS

Interpretation of findings is limited by several features of the study, in particular measurement. First and most important is the lack of an oral measure of students' knowledge of word meanings taught in the interventions. The proximal measure of reading vocabulary is confounded with student decoding ability, and therefore advantages students who received Connections. Reliability of the measure at pretest was low, reflecting limited alphabet and decoding skills in October of kindergarten. The reading vocabulary test measures how well students learn to read or at least recognize by sight the taught word form that matches a meaning read-aloud by the tester. Being able to read/decode taught vocabulary words is of value in reading school texts, and draws on both orthographic and semantic word knowledge. However, an oral test of taught-word meanings is needed to determine children's learning in the IBR condition. Without minimizing this limitation, it draws attention to a larger question: how to reliably assess vocabulary knowledge in young children who are often unable to verbally demonstrate their understanding of word meanings (see Christ, 2011; Christ, Chu, Currie, & Cipielewski, 2014; Hoffman, Teale, & Paciga, 2014; Pearson, Hiebert, & Kamil, 2007). A broader challenge for studies like this one is assessment of vocabulary knowledge in young children, even more challenging with young English learners. In this specific treatment comparison, testing only children's knowledge of reading vocabulary seriously limits claims we can make about semantic learning. However, beyond this study, we wonder if assessment of vocabulary knowledge depth in young English learner children who are learning to read might incorporate knowledge of word form as well as word meaning. As we note below, the word corpus (i.e., decodable but often difficult-to-explain words such as *act*, *on*, *off*), and the age and language skills of children in this study were challenges for measuring semantic depth, including correct syntax.

A second limitation is the lack of information on children's proficiency in their home language and their background knowledge, which may influence vocabulary learning. Knowledge of word concept in L1 influences word learning in L2, although the high-frequency word corpus makes it more likely children knew many of the taught word concepts. Third, although random assignment was employed, groups differed at pretest on decoding and spelling. To boost confidence in the findings, we controlled for these differences and focused on gain scores rather than posttest levels. A fourth limitation is that the interpretation of the meaningful gains in general receptive vocabulary by students in both conditions is limited by lack of a no-treatment control. Nevertheless, we note that we used norm-referenced scores to boost confidence that the nonzero gains were not attributable to normal expected child development.

Other limitations concern intervention content, including our assumption that Connections would build on a firm alphabet knowledge foundation at kindergarten entry. For many children in both cohorts, alphabet knowledge (letter names as well as letter sounds) was very limited and slow to develop. Limited incidental alphabet instruction was provided during the decoding practice but clearly this was not sufficient for many students. This limited alphabet foundation may have diminished potential for building the semantic-orthographic-phonological connections that depend upon alphabet knowledge. Although we did not measure alphabet knowledge at kindergarten entry, others have found that as with native English speakers, English letter-naming fluency at kindergarten entry also predicts first-grade English reading skills for English language learners (Roberts, 2005; Yesil-Dagli, 2011). English learners with more limited preschool experience and early print exposure (Capps et al., 2005) may benefit from intense early kindergarten alphabet instruction to provide a base for learning word forms and meanings (we note that we did not have access to preschool enrollment information or home story-reading practices). Another instructional limitation we recognized from our observations of the interventions concerns the word corpus taught in both treatments. We chose high-frequency root words that were also highly decodable to allow teaching both meaning and beginning decoding and spelling of words, matched to the entry-level skills of kindergarteners. This allowed testing the explicit instruction of semantic and lexical connections (in Connections), and strengthening these connections may account for the reading vocabulary advantage for taught words for Connections. Yet these words are not optimal candidates for instruction in a storybook intervention, even for kindergarten English learners with limited vocabulary. As others note, not all words are equal candidates for instruction nor do they require the same intensity of instruction (Beck & McKeown, 2007; Graves, 2000).

The findings and our observations of children's learning suggest directions for research. Both intervention approaches may be improved with less phonetically constrained word selection. The potential value of print exposure and decoding opportunities should be examined in a storybook approach using a big book or e-book format that allows calling explicit attention to the printed vocabulary words being taught, and affording brief practice blending and spelling the words. Connections children made noticeable progress in these phonics skills and children in IBR, by mid-year, were often curious and motivated to see the printed words they were being read and taught, and to try to decode them. Even brief attention to the printed word spellings enhances vocabulary learning and may be of particular value for English learners to support pronunciation (Ehri & Rosenthal, 2007). A stronger comparison of a storybook and explicit approach would match print exposures, and would include an oral measure of semantic and orthographic learning of taught words, scoring each to measure depth of learning.

CONCLUSION

Early interventions appear most promising to narrow the gap in English vocabulary knowledge for young English learners. It seems clear that for many children, the gap is so wide that no single intervention will be sufficient, and that instructional features should be evidence based, "well specified and aligned to specific risk factors" (Marulis & Neuman, 2013, p. 253). In this efficacy study we found that supplemental explicit intervention designed to build semantic, orthographic, and phonological associations for taught words allowed children to make significant growth in proximal reading vocabulary and decoding. This explicit instruction was of most benefit to children entering kindergarten with limited alphabet and decoding skills. Findings suggest future study of these instructional features for teaching a broader, less-constrained word corpus, and for teaching vocabulary to older students with established word-reading skills. Findings for the book-reading comparison treatment add to the larger body of evidence that supports explicit vocabulary instruction in the context of storybook reading for young children. Earlier research on these interventions used with preschool children has recommended the value of calling attention to print during storybook reading (Justice & Ezell, 2002; Justice et al., 2009). The present results raise the question whether vocabulary learning in the context of storybook reading may be further enhanced for primary-grade children by including brief opportunities for decoding, pronouncing, and spelling new words encountered in the stories. Both interventions were successfully delivered by teaching assistants, and these learning details warrant testing in classroom and home interventions. The magnitude of English vocabulary learning gains necessary to catch up many English learners and children at risk for academic underachievement may require strong coordination of classroom, supplemental, and home interventions that share easily implemented practices associated with strong word-learning outcomes.

REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: The MIT Press.

Adams, M. J., & Huggins, A. W. F. (1985). The growth of children's sight vocabulary: A quick test with educational and theoretical implications. *Reading Research Quarterly*, *20*, 262–281.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English learners. *Learning Disabilities Research and Practice*, *20*, 50–57.

Aukrust, V. G., & Rydland, V. (2011). Preschool classroom conversations as long-term resources for second language and literacy acquisition. *Journal of Applied Developmental Psychology*, *32*, 198–207.

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

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology*, *35*, 126–140.

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

Beck, I. L., 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.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*, 525–531.

Biemiller, A. (1999). *Language and reading success* (From Reading Research to Practice, Vol. 5). Cambridge, MA: Brookline Books.

Biemiller, A. (2003, April). *Teaching vocabulary to kindergarten to grade two children*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

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., & 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.

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.

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J.,. . .Hulme, C. (2007). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *The Journal of Child Psychology and Psychiatry*, *49*, 422–432.

Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read. A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, *65*, 1–21.

Capps, R., Fix, M. E., Murray, J., Ost, J., Passel, J., & Herwantoro-Hernandez, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington, DC: The Urban Institute.

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge, MA: Brookline Books.

Christ, T. (2011). Moving past "right" or "wrong" toward a continuum of young children's semantic knowledge. *Journal of Literacy Research*, *43*, 130–158.

Christ, T., Chu, M. M., Currie, A., & Cipielewski, J. (2014). The relation between test formats and kindergarteners' expressions of vocabulary knowledge. *Reading Psychology*, *35*, 499–528.

Christian, K., Morrison, F. J., Frazier, J. A., & Massetti, G. (2000). Specificity in the nature and timing of cognitive growth in kindergarten and first grade. *Journal of Cognition and Development*, *1*, 429–448. doi:10.1207/S15327647jCD0104_04

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, *21*, 1106–1116.

Coady, J., Carrell, P., & Nation, P. (1985). *The teaching of vocabulary in ESL from the perspective of schema theory*. Milwaukee, WI: Midwest TESOL.

Coady, J., Magoto, J., Hubbard, P., Graney, J, & Mokhtari, K. (1993). High frequency vocabulary and reading proficiency in ESL readers. *Second Language Reading and Vocabulary Learning*, *1*, 217–228.

Collins, M. F. (2010). ELL preschoolers' English vocabulary acquisition from storybook reading. *Early Childhood Research Quarterly*, *25*, 84–97.

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

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

CTB/McGraw-Hill. (2006). *Washington English Language Proficiency Assessment (WELPA)*. Monterey, CA: Author.

Cunningham, A. E., & Stanovich, K. E. (1993). Where does knowledge come from? Specific association between print exposure and information acquisition. *Journal of Educational Psychology*, *85*, 211–229.

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

Dunn, L. M., & Dunn, L. M. (2006). *The Peabody Picture Vocabulary Test: IIIA Edition*. Circle Pines, MN: AGS.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Erlbaum.

Ehri, L. C., & Rosenthal, J. (2007). Spellings of words: A neglected facilitator of vocabulary learning. *Journal of Literacy Research*, *39*, 389–409.

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

Ewers, C. A., & Brownson, S. M. (1999). Kindergarten vocabulary acquisition as a function of active vs. passive storybook reading, prior vocabulary, and working memory. *Journal of Reading Psychology*, *20*, 11–20.

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge differences by class and race. *Social Science Research*, *33*, 464–497.

Graves, M. F. (2000). A vocabulary program to complement and bolster a middle-grade comprehension program. In B. M. Taylor, M. F. Graves, & P. van den Broek (Eds.), *Reading for meaning:*

Fostering comprehension in the middle grades (pp. 116–135). New York: Teachers College Press.

Graves, M. F. (2006). *The vocabulary book*. New York, NY: Teachers College Press.

Graves, M. F., Ryder, R. J., Slater, W. H., & Calfee, R. C. (1987). The relationship between word frequency and reading vocabulary using six metrics of frequency. *Journal of Educational Research*, *81*, 81–91.

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

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap. *American Educator*, *27*, 4–9.

Hilte, M., & Reitsma, P. (2011). Activating the meaning of a word facilitates the integration of orthography: Evidence from spelling exercises in beginning spellers. *Journal of Research in Reading*, *34*, 333–345.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*, 1368–1378.

Hoffman, J. L., Teale, W. H., & Paciga, K. A. (2014). Assessing vocabulary learning in early childhood. *Journal of Early Childhood Literacy*, *14*(4), 459–481.

Juel, C. (1980). Comparison of word identification strategies with varying context, word type, and reader skill. *Reading Research Quarterly*, *15*, 358–375.

Justice, L. M., & Ezell, H. K. (2002). Use of storybook reading to increase print awareness in at-risk children. *American Journal of Speech-Language Pathology*, *11*, 17–29.

Justice, L. M., Kaderavek, J. N., Fan, X., Sofka, A., & Hunt, A. (2009). Accelerating preschoolers' early literacy development through classroom-based teacher-child storybook reading and explicit print referencing. *Language, Speech, and Hearing Services in Schools*, *40*, 67–85.

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools*, *36*, 17–32.

Kahn-Horwitz, J., Schwartz, M., & Share, D. (2011). Acquiring the complex English orthography: A triliteracy advantage? *Journal of Research in Reading*, *34*, 136–156.

Kieffer, M. J., & Vukovic, R. K. (2013). Growth in reading-related skills of language minority learners and their classmates: More evidence for early identification and intervention. *Reading and Writing: An Interdisciplinary Journal*, *26*, 1159–1194.

Kim, Y-S., Al Otaiba, S., Puranik, C., Folsom, J. S., & Gruelich, L. (2014). The contributions of vocabulary and letter writing automaticity to word reading and spelling for kindergarteners. *Reading and Writing: An Interdisciplinary Journal*, *27*, 237–253.

Lervag, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *The Journal of Child Psychology and Psychiatry*, *51*, 612–620.

Lesaux, N. K., Koda, K., Siegel, L., & Shanahan, T. (2006). Development of literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 75–122). Mahwah, NJ: Lawrence Erlbaum Associates.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology*, *95*, 482–494.

Loftus, S. M., Coyne, M. D., McCoach, D. 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 & Practice*, *25*, 124–136.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Education Research*, *80*, 300–335.

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.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition (*Vol. *2)*. Cambridge, MA: MIT Press.

McKeown, M. G., & Beck, I. L. (2014). Effects of vocabulary instruction on measures of language processing: comparing two approaches. *Early Childhood Research Quarterly, 29*, 520–530.

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

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, *20*, 522–535.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*, 267–296.

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*, 979–1007. doi:10.3102/0034654309332561

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., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*, 233–253.

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

National Center for Education Statistics (2012). *The Nation's Report Card: Vocabulary results from the 2009 and 2011 NAEP Reading Assessments (*NCES *2013 452)*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Nelson, J. R., Balass, M., & Perfetti, C.A. (2005). Differences between written and spoken input in learning new words. *Written Language and Literacy*, *8*, 101–120.

Nelson, J. R., & Vadasy, P. (2007). *Early vocabulary connections: First words to know and decode*. Longmont, CO: Sopris West.

Nelson, J. R., Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of a Tier 2 supplemental root word vocabulary and decoding intervention with kindergarten Spanish-speaking English learners. *Journal of Literacy Research*, *43*, *2*, 184–211.

Oller, D. K., & Pearson, B. Z. (2002). Assessing the effects of bilingualism: A background. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 3–21). Clevedon, UK: Multilingual Matters.

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

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

Proctor, C. P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology*, *97*, 246–256.

Ricketts, J., Bishop, D. V. M., & Nation, K. (2008). Investigating orthographic and semantic aspects of word learning in poor comprehenders. *Journal of Research in Reading*, *31*, 117–135.

Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, *11*, 235–257.

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

Roberts, T. A. (2005). Articulation accuracy and vocabulary size contributions to phonemic awareness and word reading in English language learners. *Journal of Educational Psychology*, *97*, 601–616.

Roberts, T. A., & Neal, H. (2004). Relationships among preschool English language learners' oral proficiency in English instructional experience and literacy development. *Contemporary Educational Psychology*, *29*, 283–311.

Rosenthal, J., & Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology, 100*, 175–191.

Rosenthal, J., & Ehri, L. C. (2011). Pronouncing new words aloud during the silent reading of text enhances fifth graders' memory for vocabulary words and their spellings. *Reading and Writing*, *24*, 921–950.

Roskos, K., Ergul, C., Bryan, T., Burstein, K., Christie, J., & Han, M. (2008). Who's learning what words and how fast? Preschoolers' vocabulary growth in an early literacy program. *Journal of Research in Childhood Education*, *22*, 275–290.

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.

Schwartz, M., Kahn-Horwitz, J., & Share, D. (2014). Orthographic learning and self-teaching in a bilingual and biliterate context. *Journal of Experimental Child Psychology*, *117*, 45–58.

Senechal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language*, *24*, 123–138.

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

Silverman, R. (2007). A comparison of three methods of vocabulary instruction during read-alouds in kindergarten. *The Elementary School Journal*, *108*, 97–113.

Silverman, R., & Crandell, J. D. (2010). Vocabulary practices in prekindergarten and kindergarten classrooms. *Reading Research Quarterly*, *45*, 318–340.

Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English language learners in pre-kindergarten through second grade. *Journal of Educational Psychology*, *101*, 305–314.

Skibbe, L. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2011). Schooling effects on preschoolers' self-regulation, early literacy, and language growth. *Early Childhood Research Quarterly*, *26*, 42–49.

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

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

Swanson, H. L., Rosston, K., Gerber, M., & Solari, E. (2008). Influence of oral language and phonological awareness on ELL children's bilingual reading. *Journal of School Psychology*, *46*, 413–429. doi:10.1016/j.jsp.2007.07.002

Swanson, H. L., Saez, L., & Gerber, M. (2006). Growth in literacy and cognition in bilingual children at risk for reading disabilities. *Journal of Educational Psychology*, *98*, 247–264.

Vadasy, P. F., Nelson, J. R., & Sanders, E. A. (2013). Longer-term effects of a Tier 2 kindergarten vocabulary intervention for English learners. *Remedial and Special Education*, *34*, 27–37.

Verhoeven, L. (2011). Second language reading acquisition. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research* (Vol. IV) (pp. 661–683). New York, NY: Routledge.

Wasik, B., & Bond, M. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology*, *93*, 243–250.

Wasik, B., Bond, M., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology*, *98*, 63–74.

What Works Clearinghouse. (2014, March). *Procedures and standards handbook, version 3.0.* U.S. Department of Education, Institute of Education Sciences.

Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Schultz, M. D., Velting, O. N., & Fischel, J. E. (1999). Outcomes of an emergent literacy intervention from Head Start through second grade. *Journal of Educational Psychology*, *91*, 261–272.

Wilkinson, G. S., & Robertson, G. J. (2006). *The Wide Range Achievement Test-4*. Lutz, FL: *Psychological Assessment Resources*.

Woodcock, R. W. (1987/1998). *Woodcock Reading Mastery Tests-Revised/Normative Update (WRMT-R/NU)*. Circle Pines, MN: American Guidance Service.

Wright, T. S. (2012). What classroom observations reveal about oral vocabulary instruction in kindergarten. *Reading Research Quarterly*, *47*, 353–355.

Wright, T. S., & Neuman, S. B. (2013). Vocabulary instruction in commonly used kindergarten core curricula. *The Elementary School Journal*, *113*, 386–408.

Yesil-Dagli, U. (2011). Predicting ELL students' beginning first-grade English oral reading fluency from initial kindergarten vocabulary, letter naming, and phonological awareness skills. *Early Childhood Research Quarterly*, *26*, 15–29.

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

Zipoli, R. P., Coyne, M. D., & McCoach, D. B. (2011). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically related and embedded word review. *Remedial and Special Education*, *32*, 131–143.

APPENDIX

Table A1. Target words and meanings taught in interventions

ant: An ant is a small insect with six legs that lives in nests.	**pit:** A pit is a big hole in the ground.
cap: A cap is something you wear on your head.	**rib:** A rib is a curved bone that is connected to your spine, or back.
cat: A cat is an animal that purrs and has whiskers and a long tail.	**sip:** A sip is a very small taste of a drink.
mop: You use a mop to clean the floor.	**sit:** To sit means to rest on your bottom.
pan: You use a pan to cook or bake food.	**thin:** Thin means skinny or narrow.
add: To add means to put numbers or things together.	**tin:** Tin is a kind of metal.
mad: To be mad is to feel angry.	**with:** With means to be together or to be a part of something.
map: A map is a drawing you use to find a place.	**dash:** To dash means to move very fast.
on: On means at the top of something or touching something.	**ship:** A ship is a big boat that carries people and things.
act: Act is how you behave or seem.	**shop:** To shop means to look for things we want to buy.
damp: Something that is damp is wet or moist.	**wish:** A wish is something you hope for or want to happen.
dot: A dot is a small, round spot.	**wax:** Wax means a liquid or a paste you rub on things to make them shine.
off: Off means to remove or stop something.	**win:** To win means to do the best.
nap: A nap is a short sleep during the day.	**wing:** A wing is a part of a bird or a plane that helps it fly.
pants: Pants are clothes that cover your legs from the waist to the ankle.	**jam:** To jam means to push and cram things together.
nod: To nod means to move your head up and down.	**job:** A job is work you do to earn money.
pass: To pass means to go by or get in front of someone or something.	**bud:** A bud is a flower or a leaf that has not opened yet.
pack: To pack means to put things into a bag or a box.	**bump:** To bump means to hit or run into someone or something.

(Continued on next page)

Table A1. *(Continued)*

sack: A sack is a bag that is used to hold or carry things.	**cub:** A cub is a baby bear, lion, or fox.
hop: The word hop means to jump up and down.	**hut:** A hut is a small, simple house.
hot: Something that is hot is very warm.	**mud:** Mud is wet, sticky dirt.
gas: Gas is what we use to run our cars and heat our houses.	**just:** Just means only.
tag: A tag is a small sign that tells you something.	**must:** The word must means that you have to do something.
rat: A rat is an animal with a long tail that looks like a mouse, only bigger.	**rush:** To rush is to do something very quickly.
rock: A rock is a big stone.	**tug:** To tug means to pull hard.
band: A band is a thin, flat strip of rubber or material that stretches.	**bank:** A bank is a safe place to keep money.
bat: A bat is a long stick with a round tip that is used to hit things.	**kid:** A kid is a baby goat.
back: The back is the rear part of something or of your body.	**think:** To think means to use your brain to consider or study something.
boss: A boss is the person in charge at work.	**wink:** To wink is to quickly close and open one eye.
bath: When you take a bath, you wash your body in a tub of water.	**ill:** If you are ill you are sick or not feeling well.
path: A path is a trail for walking.	**log:** A log is a big, long piece of wood from a tree.
big: Big means very large or tall.	**long:** Long means that something will take a lot of time to do or finish.
dig: Dig is to turn over or remove dirt or sand.	**land:** Land is soil, earth, or dirt.
dim: Dim means not bright or shiny.	**last:** The word last means at the end or the only thing or person left.
grin: A grin is a smile that shows you are happy. To grin is to smile.	**list:** A list is a set of things you write on paper.
hid: If you hid something, you put it where it can't be seen.	**luck:** The word luck means something very good that happens.
hit: To hit means to strike or punch something.	**best:** The word best means better than all others.
in: The word in is short for the word inside.	**mess:** If something is a mess, it means that things are out of place or dirty.
pick: To pick means to choose something or someone.	**pet:** A pet is an animal that you keep at home.
pin: A pin is a small piece of wire with a sharp end that is used to hold things together.	**rest:** To rest means to not do anything for awhile.

Table A2. Primary instructional attributes of experimental conditions

Attribute	Connections	IBR
Target words	• **High-frequency decodable root words explicitly defined** • Mean number of word repetitions in lesson = 8	• **High-frequency decodable root words explicitly defined** • Mean number of word repetitions in lesson = 3
Related words	• Synonyms, antonyms, related words (varied depending on target word: student-generated with tutor support)	• Conceptually connected words featured in the storybook (2–4: tutor-provided and -defined)
Word meanings	• **General approach: Oral interactions with target word** • **Oral presentation of child-friendly definitions supported with pictures** • **Independent oral use of the words in child-generated sentences** • Student reading of decodable passage using the word in context with appropriate tutor support • Use of sentence context (cloze procedure) to identify appropriate word (from 3–4 choices) including target word • Compare/contrast attributes of pictures depicting example and nonexamples of target word meaning	• **General approach: Oral interactions with target word** • **Oral presentation of child-friendly definitions supported with pictures** • **Independent oral use of the words in child-generated sentences** • Tutor reading of book to children and referencing the meaning of target words when encountered and discussion using scripted questions specific to each story
Decoding	• Spelling and word blending with cumulative review with appropriate tutor support	

Note. Items in boldface indicate similarities in instruction among conditions.

Table A3. Instructional activities for connections treatment

1. ***Word blending and spelling.*** The tutor modeled decoding and spelling aloud the target word, and students practiced oral decoding and spelling. Target words were presented randomly (three times) within a serial list that also featured two recently taught words. After reading (decoding and pronouncing) the list of words, the students reread and spelled aloud the target word. The printed target word was visible to the student throughout the lesson.
2. ***Word meaning.*** The tutor read a student-friendly definition for the target word. The meaning for each target word was depicted by a picture and the lesson included a sentence to support the definition. For example, the definition and sentence provided for the word "nap" was:

   A ***nap*** *is a short sleep during the day.*
   My baby sister takes a ***nap*** *every afternoon.*

   After explaining the meaning and using the picture and sentence to tell about the word, the tutor asked the student to tell the meaning. The tutor prompted students to talk about the meaning for the target word, asking students to think about other words related to the target word (synonyms, antonyms, word associations), and evaluating students' responses and expanding or correcting students' responses by rephrasing and adding information.
3. ***Fast read passage.*** Students read a fully decodable short passage that featured the target word in a context written to be meaningful to young children. While most of the passage was written to be decodable based on the letter sound sequence used in the program, the last sentence in each passage was a less decodable sentence meant for the teacher to read. This last sentence reinstated the meaning of the target word using the context of the passage. For example, the passage for the word "nap" consisted of three decodable sentences and one less decodable sentence the teacher and students read together:

   Cat had a nap.
   Cat had a nap on Nan's cap.
   A cat had a nap on a cap.
   The cat did not sleep long—he just had a nap.

   Each passage was first read aloud together by the teacher and student. The teacher–student read-aloud was then followed by a student read-aloud, with scaffolding as needed. Each read-aloud was repeated until students could read the passage fluently and show comprehension. Students were asked to think about the meaning for the target words prior to each read-aloud. The teacher checked for understanding at the end of each read-aloud.
4. ***Sentence completion***. Students completed a "fill-in-the-blank" activity in a written cloze procedure task that provided scaffolded syntactic awareness practice. Students chose from three printed word choices. For example:

Sentence	**Word Choices**
Cat had a nap.	
Cat had a ———— on Nan's cap.	*cap*
A cat had a nap on a ————!	*nap*
The ———— didn't sleep long—he just had a nap.	*cat*

5. ***Word meaning match***. Students had to identify the picture for the target word from a pair of pictures illustrating an example and nonexample of the meaning. The tutor helped the student describe why (or why not) the picture illustrated the target word.

(Continued on next page)

Table A3. *(Continued)*

6. ***Say a sentence***. The lesson ended with a production task that required students to use the word in a sentence. Unique prompts for each word were provided so that tutors did not have to create an appropriate and engaging prompt. Prompts varied in the degree to which they challenged students, and allowed the tutor to individualize instruction for students within a group. Some prompts were easier and required students to discriminate between two words (e.g., "Tell me how a *bud* is different from a *flower*"), and other prompts challenged students to use the correct word in context (e.g., "Tell me what it is like if the light is *dim*").

Table A4. Instructional procedures for IBR treatment

1. The tutor introduced the definition of the target word *before* reading the story. The student was shown a card with a picture of the word on one side and the printed word and definition on the opposite side of the card (seen by tutor only). After telling students the word and meaning, the tutor asked students to tell the meaning, to tell what they knew about the word, and to use the word in a sentence.
2. The tutor introduced 2–4 words related in some way to the target word that also appeared in the story. These additional words often developed semantic relationships with the target word. For example, for the target word *damp*, related story words were *hose* and *towel*. For the target word *win*, related words were *game* and *play*. The tutor used the same procedures described above for the target word to introduce and engage in conversation about the related words.
3. The tutor introduced the story (title, cover, illustration) and if appropriate asked the student to predict what the story was about, drawing on background knowledge.
4. The tutor read aloud the story, stopping every few pages to ask both recall and open-ended questions about the story, and questions connected to the target word when it appeared. Tutors were provided with scripted questions for each book and were encouraged to adapt questions to the changing skill levels and interests of students.
5. After reading the story, the tutor reviewed the word cards for the story, including the definition for the target and connected words. The tutor asked students to tell the meaning of the target word in their own words, and to use the target word in a sentence. Similar to the *Connections* instruction, tutors were trained to help students expand on their responses. Early in the year tutors helped students add 1–2 words to single-word responses. As students' language skills developed, tutors helped students to respond in complete sentences and to expand in more complex sentence responses.

Table A5. Fidelity components for experimental conditions

Fidelity Component Criteria	CONN	IBR
Target words		
Tutor presents the target word	X	X
Tutor modeled blending of the target word	X	
Students blended the target word	X	
Students blended the specified practice words	X	
Students spelled the target word	X	
Tutor presented the specified word meaning	X	X
Tutor used the picture to support his or her explanation		X
Tutor used the specified sentence to support his or her explanation	X	
Students generated the meaning/used word in sentence	X	X
Tutor evaluated, rephrased, and expanded student responses	X	X
Related words		
Tutor presented the specified connected words		X
Tutor evaluated, rephrased, and expanded on student responses		X
Text reading		
Tutor reads the text	X	X
Tutor and students read the text	X	
Tutor referenced the target word when it was encountered in the text		X
Tutor asked questions, prompted student responses during reading of the story		X
Tutor used interactive approach to identify the correct missing words in the sentence completion text	X	
Tutor evaluated, rephrased, and expanded on student responses	X	X
Word review		
Tutor asked students to identify pictures that represented the meaning of the target word	X	
Tutor asked students to explain their picture choice	X	
Tutor asked students meaning for target word and use in a sentence	X	X
Tutor evaluated, rephrased, and expanded on student responses	X	X

Note. Items in boldface indicate shared (common) fidelity components among conditions.

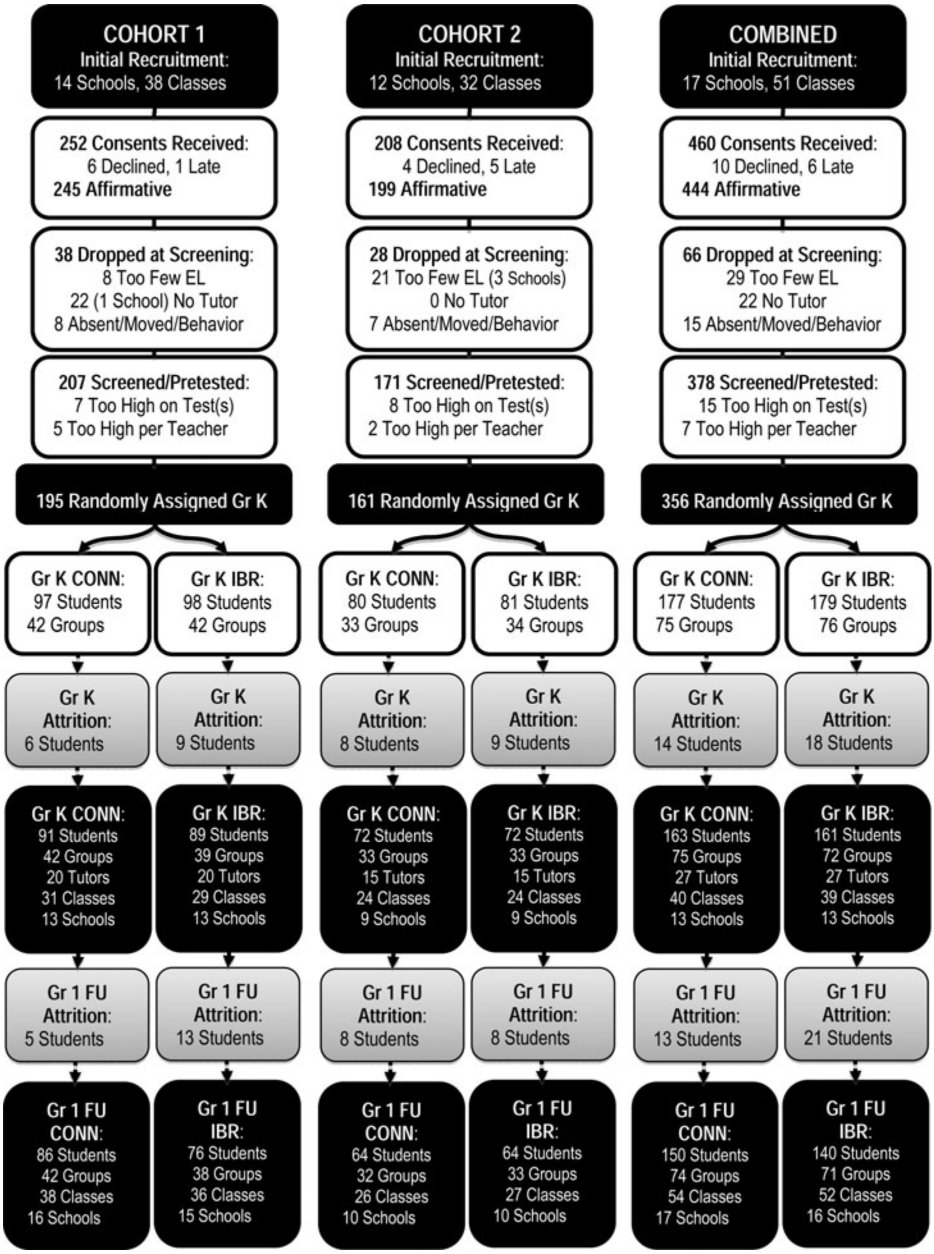

Figure A1. Flowchart of recruitment, assignment, and attrition by cohort. *Note*. Combined cohorts' schools, classrooms, and tutors do not directly sum due to overlap across cohorts.

APPENDIX B

Table B1. Disaggregated zero-order correlations by experimental condition

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
Treatment-related														
1. Attendance	—	0.09	−0.10	−0.07	0.10	0.02	**0.20**	0.10	0.11	0.12	**0.19**	0.01	0.11	0.07
2. Fidelity	0.11	—	0.14	0.10	**0.15**	0.05	0.08	**0.19**	0.09	**0.20**	−0.03	0.08	0.12	0.14
Pretests														
3. Receptive vocabulary	**0.14**	0.05	—	0.12	0.02	**0.27**	**−0.60**	**0.26**	**0.25**	−0.07	**−0.71**	**0.31**	**0.20**	0.08
4. Reading vocabulary	−0.01	0.02	**0.26**	—	−0.01	−0.06	−0.14	**−0.38**	−0.05	−0.04	**−0.17**	**−0.43**	0.02	0.05
5. Decoding	**0.17**	−0.07	**0.23**	0.10	—	0.10	−0.03	**0.19**	−0.08	−0.02	−0.07	0.11	−0.13	−0.03
6. Spelling	**0.29**	0.04	**0.54**	0.13	**0.38**	—	−0.04	**0.45**	**0.51**	**−0.44**	0.02	**0.41**	**0.46**	**−0.37**
Immediate gains														
7. Receptive vocabulary	0.02	−0.07	**−0.49**	**−0.16**	0.05	−0.04	—	0.13	0.09	**0.22**	**0.72**	0.06	0.08	0.11
8. Reading vocabulary	**0.34**	0.03	**0.37**	**−0.26**	**0.33**	**0.58**	0.03	—	**0.50**	0.12	0.05	**0.68**	**0.38**	0.11
9. Decoding	**0.18**	0.06	**0.36**	0.13	−0.04	**0.48**	0.03	**0.49**	—	**0.18**	0.07	**0.52**	**0.61**	**0.25**
10. Spelling	−0.09	0.08	−0.13	−0.09	−0.02	**−0.38**	**0.15**	0.03	0.12	—	0.08	0.11	0.12	**0.63**
Longer-term gains														
11. Receptive vocabulary	0.02	−0.03	**−0.57**	**−0.28**	0.12	−0.04	**0.69**	0.09	−0.09	**0.18**	—	0.05	0.09	0.03
12. Reading vocabulary	**0.24**	0.05	**0.30**	**−0.36**	**0.31**	**0.51**	0.10	**0.77**	**0.47**	0.14	0.15	—	**0.46**	**0.21**
13. Decoding	**0.18**	0.15	**0.27**	0.05	−0.07	**0.41**	0.11	**0.44**	**0.68**	0.16	−0.02	**0.47**	—	**0.31**
14. Spelling	−0.05	0.08	−0.07	−0.12	0.01	**−0.33**	0.08	0.11	**0.19**	**0.58**	**0.17**	**0.24**	**0.33**	—

Note. Connections condition ($N = 163$ students at pretest and posttest, 150 at follow-up) given in lower diagonal, IBR condition ($N = 161$ students at pretest and posttest, 140 at follow-up) given in upper diagonal. Attendance = number of days student attended tutoring; Fidelity = adjusted average (model-implied) small-group common components fidelity % across intervention period; Receptive vocabulary = standard score of Peabody Picture Vocabulary Test-IIIA; Reading Vocabulary = percent correct out of 25 items on curriculum-based measure of root word vocabulary; Decoding = standard score of Woodcock Reading Mastery Test-Revised/Normative Update Word Attack subtest; Spelling = standard score of Wide Range Achievement Test-4 Spelling subtest. Pretest assessed in fall of kindergarten; Immediate gains = posttest (spring of kindergarten)—pretest; Longer Term Gains = follow-up (winter of first grade)—pretest. Pearson's r shown for disaggregated student-level data; correlations in boldface are significant at the .05 level.