

# The Effectiveness and Ease of Implementation of an Academic Vocabulary Intervention for Linguistically Diverse Students in Urban Middle Schools

Nonie K. Lesaux

*Harvard University, Cambridge, Massachusetts, USA*

Michael J. Kieffer

*Columbia University, New York, USA*

S. Elisabeth Faller, Joan G. Kelley

*Harvard University, Cambridge, Massachusetts, USA*

## ABSTRACT

The present study aims to advance the extant research base by evaluating the implementation and effectiveness of an academic vocabulary program designed for use in mainstream middle school classrooms with high proportions of language minority learners. The quasi-experimental, mixed-methods study was conducted in 21 classes (13 treatment matched to 8 control) in seven middle schools in a large district, with 476 sixth-grade students (346 language minority learners, 130 native English speakers). Classroom observations and teacher logs indicated the 18-week program was implemented with good fidelity and that the approach contrasted sharply with the standard district English language arts (ELA) curriculum. Multilevel modeling indicated that the program resulted in significant effects on several aspects of vocabulary knowledge, including meanings of taught words ( $d = 0.39$ ;  $p < .0001$ ), morphological awareness ( $d = 0.20$ ;  $p = .0003$ ), and the word meanings as presented in expository text ( $d = 0.20$ ;  $p = .0227$ ). The program also yielded marginally significant, but promising effects on a depth of word knowledge measure ( $d = 0.15$ ;  $p = 0.0830$ ) and a norm-referenced measure of reading comprehension ( $d = 0.15$ ;  $p = .0568$ ). No effects were found on a norm-referenced vocabulary measure. These effects were comparable for language minority learners and their native-English-speaking classmates. Data from teachers shed light on the challenges of meeting students' diverse instructional needs and the roles of curriculum and professional networks in building instructional capacity. The findings show promise in developing effective multifaceted vocabulary instruction for implementation by ELA teachers in middle school classrooms with high numbers of language minority learners.

Students attending schools in urban settings rarely experience the same opportunities to learn literacy skills as their counterparts in suburban settings (Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Hakuta, 1998; Snow, Burns, & Griffin, 1998). Urban schools—schools located in large city centers and characterized by high concentrations of students of color and students from low-income backgrounds—have

historically suffered from limited educational resources (e.g., shortages of qualified teachers, high teacher turnover, large class sizes, inadequately rigorous curriculum; Anyon, 1980, 1997; Lankford, Loeb, & Wyckoff, 2002; National Center for Education Statistics (NCES), 2009a, 2009b). It is thus not surprising that students in urban schools disproportionately demonstrate below-average outcomes on large-scale assessments. For

example, a recent analysis of National Assessment of Educational Progress (NAEP) data from large city districts demonstrated that 10 of the 11 participating districts had high—in some cases staggeringly high—proportions of learners scoring below established proficiency levels (Lutkus, Grigg, & Donahue, 2007). In 6 of the 11 districts studied, more than 50% of students scored *below basic* in reading as compared with the national rate of 34%. These differences in student outcomes by school context raise questions about what types of instructional and curricular resources can be brought to bear to improve reading performance in settings characterized by low performance; increasing opportunities to learn for students in urban schools is imperative for their personal and academic success. Of particular concern is the growing population of language minority learners, who often have even fewer opportunities to learn than their native-English-speaking peers within the same schools (Gándara et al., 2003; Hakuta, 1998; Snow et al., 1998); a large proportion of these learners demonstrate reading comprehension difficulties, particularly after the primary grades (August & Shanahan, 2006).

A step toward effective reform to increase adolescent literacy rates is the development of empirically based approaches to promote students' reading comprehension skills (Biancarosa & Snow, 2006; Moore, Bean, Birdyshaw, & Rycik, 1999). Limited vocabulary knowledge is a potential source of reading comprehension difficulties, especially among older struggling readers, whether language minority learners or native English speakers (e.g., Bailey, 2006; Biancarosa & Snow, 2006; Chall & Jacobs, 2003; Fillmore, 1982; National Institute of Child Health and Human Development [NICHD], 2000; RAND Reading Study Group, 2002; Stahl & Nagy, 2006; Valdés, 2000). The relationship between vocabulary and reading comprehension is thought to be reciprocal—knowing more words facilitates successful comprehension, while successful comprehension and wider reading lead to opportunities to learn more words (Freebody & Anderson, 1983; Stanovich, 1986). Differences in both vocabulary knowledge and reading outcomes between native speakers and language minority learners have been shown to widen over time (Kieffer, 2008; Nakamoto, Lindsey, & Manis, 2007). Thus, the primary goal of the present study was to evaluate the effectiveness of an academic vocabulary program designed to promote the reading comprehension skills of learners in low-performing urban middle school classrooms with high numbers of language minority learners. A secondary goal was to investigate its ease of implementation, defined by teachers' use and their insights about the program implementation.

---

## Vocabulary Development

Reading comprehension is a complex skill that requires higher level processing, such as drawing on prior knowledge, making inferences, and resolving structural and semantic ambiguities while reading, and also involves the integration of many specific linguistic and cognitive skills (e.g., word reading, syntactic awareness) with background and cultural knowledge (Alexander & Jetton, 2000; Kintsch, 1994; McNamara, Kintsch, Songer, & Kintsch, 1996). Each of these components of effective comprehension necessarily draws on vocabulary knowledge; once words are decoded accurately, the reader must grasp the words' meanings to comprehend clauses, propositions, and paragraphs.

Vocabulary knowledge is likewise multidimensional and complex in nature; knowing a word well requires a combination of different types of knowledge: its definition, its relationship to other words, its connotations in different contexts (i.e., polysemy), and its transformation into other morphological forms (Nagy & Scott, 2000; Stahl, 1999). Knowledge of a word—particularly an abstract, conceptually sophisticated word—is thought to develop incrementally over time, with students gaining additional information about a word with each meaningful, contextualized encounter with it. Thus, vocabulary is considered both a key mechanism in and an important effect of metacognitive processes within the reading comprehension process; students who know more words have more abstract language at their disposal with which to be strategic while reading, and students with developed understanding of language and strategies to manipulate language will learn words more successfully (RAND Reading Study Group, 2002; Ruddell & Unrau, 2004; Sweet & Snow, 2004). Whereas skilled readers acquire much of their vocabulary through encounters with unfamiliar words while reading (Sternberg, 1987), children with impoverished vocabularies cannot necessarily rely on learning words through wide reading. Because struggling readers read less than their typically achieving peers, they encounter fewer words, especially low-frequency words, than do skilled readers (Stanovich, 1986). These learners also suffer from less developed metacognitive strategies for word learning; they are less equipped to use surrounding words and grammatical clues to glean the meaning of unfamiliar words from context, and often cannot rely on surrounding known words because the ratio of known to unknown words is too high (e.g., Carver, 1994; Stoller & Grabe, 1995).

Language minority learners are especially likely to have underdeveloped English vocabularies (August, Carlo, Dressler, & Snow, 2005), even when from middle-income backgrounds (Umbel, Pearson, Fernandez, & Oller, 1992). The largest and fastest-growing

populations of language minority learners in the U.S. are students who immigrated before kindergarten and U.S.-born children of immigrants (Capps et al., 2005). This population enrolls in U.S. schools at kindergarten; by middle school, few of these learners need instruction in basic English, yet many lack the academic English vocabulary central to text comprehension and school success (Scarcella, 2003). For example, in several studies with language minority learners in the elementary and middle school years—whether formally designated limited English proficient or not—these students' vocabulary levels are well below average (Manis, Lindsey, & Bailey, 2004; McLaughlin et al., 2000; Proctor, Carlo, August, & Snow, 2005; Swanson, Saez, & Gerber, 2006). However, of significant importance when considering the design and delivery of effective intervention strategies to serve language minority learners is that comparative work suggests these vocabulary levels are only slightly lower than many of their classmates who are monolingual speakers (Beech & Keys, 1997; Droop & Verhoeven, 2003; Hutchinson, Whiteley, Smith, & Connors, 2003; Leseman & de Jong, 1998; Verhoeven, 2000). In these studies, the practical importance of this difference is minimal, suggesting the promise of a whole-classroom approach to intervention. Thus, the design of this study was based on the premise that many struggling readers in mainstream classrooms—whether language minority or not—lack the specialized academic language of text needed for school success. In analyzing program effects, therefore, a guiding question was whether there were differences between language minority learners and their monolingual classmates in their response to instruction.

---

## Vocabulary Instruction

For several decades, research in elementary schools has documented that most classrooms incorporate little, if any, systematic and explicit vocabulary instruction into the curriculum (Durkin, 1978; Roser & Juel, 1982; Scott, Jamieson-Noel, & Asselin, 2003; Watts, 1995). Although less is known about typical instruction in middle school classrooms, it is likely that they incorporate even less attention to vocabulary, given their focus on literature and content learning and given that supplemental interventions provided for struggling readers in the middle and high school grades tend to focus on word-level reading skills, particularly in the domain of fluency, with only superficial attention, if any, given to language development (Deshler, Palincsar, Biancarosa, & Nair, 2007). Going forward, research must focus on effective instructional practices to promote vocabulary development as it relates to comprehending and analyzing texts in the middle and high school years, and

simultaneously document the ways in which these practices as implemented differ from standard teaching practices in the English language arts classroom.

Three guiding principles of vocabulary instruction anchored the approach that we developed and evaluated in the present study. First, in light of the complexity of knowing a word, there is a growing consensus that vocabulary instruction should be characterized by the promotion of deep understanding of a relatively small number of words, their elements, and semantically and morphologically related words in rich contexts (e.g., Graves, 2000, 2006; Stahl & Nagy, 2006). This contrasts with the practice, sometimes found in vocabulary curricular materials, of teaching shallow knowledge of a large number of words from a list or workbook, with each word's definition taught on one or two occasions, which is not likely to influence reading comprehension skills. Second, the words to be taught deeply should be very high utility in nature; specifically, the words taught should be general-purpose academic words (e.g., *analyze*, *frequent*, *abstract*; Beck, McKeown, & Kucan, 2002; Graves, 2000, 2006; Stahl & Nagy, 2006), as opposed to the low-frequency and often relatively unimportant, if colorful and exotic, words (e.g., *refuge*, *burrowed*) that are sometimes selected by teachers or targeted for instruction by textbooks (Hiebert, 2005). Word selection is especially important when teaching students with impoverished vocabularies; given the enormity of the word-learning task, no teacher or curriculum can teach or expose students to the thousands of unknown words needed for academic success. Thus, the third principle for effective vocabulary instruction that guided the development of the program is to balance direct teaching with teaching word-learning strategies to equip students with cognitive tools to learn words independently, such as using contextual cues (Fukkink & de Glopper, 1998; Swanborn & de Glopper, 1999) and using one's morphological awareness skills (Baumann et al., 2002; Baumann, Kame'enui, & Ash, 2003; Kieffer & Lesaux, 2008; Nagy, Berninger, & Abbott, 2006).

Having identified these guiding principles, we reviewed existing research to provide us with information about the nature of the instruction itself, drawing on several reviews of research (Baumann et al., 2003; Beck et al., 2002; Graves, 2000; Stahl & Fairbanks, 1986) as well as individual empirical studies (Baumann et al., 2002, 2003; Beck & McKeown, 2007; Carlo et al., 2004; Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009; Lubliner & Smetana, 2005). Thus, there are several features of the program lessons and activities to promote student knowledge, including frequent encounters with the words across contexts; the use of definitional and contextual information to learn and determine word meanings; a balance of direct instruction in word meanings with teaching cognitive strategies for word learning;

and the crucial element of designing lessons to promote the active processing of target word meanings. In creating opportunities for active processing, we also committed to integrating the four language skills—listening, speaking, reading, and writing—to enrich students' opportunities to learn (e.g., Calderón, 2007; Peregoy & Boyle, 2008) and to ensure the program was suited for the standards-based English language arts classroom (California Department of Education, 1997). Finally, we incorporated the use of collaborative learning activities, given the strong sociolinguistic evidence for the benefits of peer interaction for language and metacognitive development (see August & Hakuta, 1997; Ellis, 1994; McLaughlin, 1985).

In spite of these principles, which may well apply to older learners, existing research on vocabulary instruction has been largely limited to students in elementary schools. For example, of the 50 vocabulary intervention studies focused on monolingual English speakers that were reviewed by The National Reading Panel (NICHD, 2000), 39 were conducted with students in elementary school. Moreover, the majority of vocabulary studies—whether conducted with young students or adolescents—are short-term training studies, with intervention often administered in small groups, and focused primarily on one aspect of vocabulary or strategy for word learning (for relevant reviews see Fukkink & de Gloppe, 1998; NICHD, 2000; Stahl & Fairbanks, 1986). Since the National Reading Panel Report in 2000, a handful of studies have been published—some investigating the effects of a comprehensive approach to vocabulary instruction (e.g., Beck & McKeown, 2007; Lubliner & Smetana, 2005; Nelson & Stage, 2007), some comparing single approaches (e.g., Coyne et al., 2009). Together, these studies vary significantly with respect to age and target population but suggest the promise of these techniques as well as the challenge of moving students' vocabularies forward.

The vocabulary intervention research conducted with language minority learners is extremely scarce and limited in scope. In their review, Shanahan and Beck (2006) identified only three vocabulary intervention studies published between 1980 and 2005 that were conducted with language minority learners (Carlo et al., 2004; Pérez, 1981; Vaughn-Shavuo, 1990). Of the three studies, two were conducted with young students (Pérez, 1981; Vaughn-Shavuo, 1990), one of which featured small-group, daily instruction (30 minutes) for 3 weeks, while the other featured whole-class, daily instruction (20 minutes) for 12 weeks.

There is an obvious need for more research focused on empirically based effective vocabulary instruction for language minority learners, particularly in the middle school grades when the curriculum demands that students have a command of many abstract and

conceptually complex words to read textbooks independently. If such research is going to inform instructional improvement at scale, it must be conducted in ecologically valid ways. First, there is a need for studies that test classroom-based interventions that feature a comprehensive approach to vocabulary instruction that can be readily adopted by teachers, in contrast to training studies that focus on only one aspect of vocabulary knowledge. Second, given the linguistic diversity inherent in today's school-aged population, interventions should be designed with language minority learners in mind, but should also be appropriate for their monolingual classmates; as noted, analyses should focus on the effects on each group of students, ensuring interventions meet the specific needs of language minority learners and do not carry hidden costs for the monolingual speaker. Finally, these studies must also focus on the important role of student engagement when designing instruction for the middle school student who struggles with reading, and must similarly focus on providing teachers with programs that can be implemented with ease.

There is only one published study to date that has evaluated the efficacy of classroom vocabulary instruction for children from linguistically diverse backgrounds who are beyond the primary grades, in this case upper elementary students. In their evaluation study, conducted with 254 fifth-grade English-language learners and their monolingual English-speaking classmates from 9 classrooms in 4 schools, Carlo et al. (2004) used a 15-week program (Vocabulary Improvement Program; Lively, August, Carlo, & Snow, 2003). The program is comprehensive in nature, focusing on a balance of direct instruction in word knowledge and word-learning strategies and incorporating text into word study. Measured against instruction in comparison schools, the program produced significant gains in vocabulary knowledge and reading comprehension on researcher-developed tasks for both language minority learners and their native-English-speaking classmates. Consistent with the current principles of effective vocabulary instruction, the authors concluded that the program's emphasis on learning *about* words over simply learning words is reflected in the results of the study. Although the results from this single study highlight the promise of these instructional principles and have some application to the present study, they require confirmation in different instructional contexts and using different materials. In particular, there is a pressing need for research in middle school contexts to determine if these principles can be integrated into effective materials that are appropriate for early adolescent learners and their teachers.

The present study aims to advance the extant research base by evaluating the effects of a program

designed to bring theoretically based strategic and explicit vocabulary instruction into low-performing middle school classrooms with high numbers of language minority learners. Developing empirically based instructional programs for use in middle schools represents a first, important step toward student improvement. A second important step is to explicitly consider the process of increasing opportunities to learn in urban middle schools, i.e., schools in large city centers characterized by high concentrations of students of color or from low-income backgrounds and possessing limited educational resources. Improving instruction is not just about designing effective programs to meet students' needs, but also about meeting teachers' needs such that the program can be implemented with relative ease and with efficacy (Donovan, Wigdor, & Snow, 2003).

Despite the importance of vocabulary knowledge for older readers and its role in the academic difficulties of many learners, it is rarely an emphasis in middle school curricula. Historically, middle schools have retained distinctions between English language arts and content courses, with English language arts curricula emphasizing literary response and analysis, rather than comprehension of a variety of texts. Moreover, many middle school teachers consider themselves teachers of academic content, not as literacy teachers (for relevant discussions see Frey, 2002; Lee, Dedrick, & Smith, 1991; Rex & Nelson, 2004). As a result, reading instruction is conceptualized as the function of primary school curricula, to "prepare" students for successful achievement in the middle school years. Although policymakers have begun to shift their attention to instructional reform in middle school literacy, instruction in vocabulary development and related comprehension skills for learning from text continues to play only a very minor role, if any, in middle schools across the nation (Biancarosa & Snow, 2006; National Council of Teachers of English, 2006; Wade & Moje, 2000). Thus, if the evidence is going to contribute to middle school literacy reform at scale, the challenge for researchers is in fact twofold: (1) to demonstrate evidence of efficacious literacy programs for mainstream middle school classrooms; and (2) to identify potential factors that may influence their ease of implementation and sustained use in mainstream classrooms.

---

## Present Study

The primary goal of the present study was to generate new insights related to vocabulary instruction by evaluating the effects of a vocabulary program designed—based on the principles delineated earlier—for use in low-performing middle schools with high numbers of

language minority learners in order to bolster students' vocabulary and reading comprehension skills. Given that the ultimate success of such a program will also depend upon its usefulness in engaging teachers and students in this challenging work, a secondary goal was to gain insights into its use and ease of implementation in these mainstream classrooms. In this light, we collected data on fidelity of implementation (i.e., amount of time, lessons completed, quality of targeted instructional elements) as well as data on teachers' insights and perceptions of implementing such an instructional approach (i.e., their views of the supports and challenges involved in improving instruction). Finally, given the backdrop of middle school literacy reform, we also collected data to shed light on how this theory-based vocabulary instruction contrasted with standard practice in the participating classrooms.

Although the study was designed to address the theoretical and practical need for vocabulary intervention research as described previously, it was also designed to have technical characteristics that meet the criteria set forth by Slavin and colleagues (most recently articulated in Slavin, Cheung, Groff, & Lake, 2008) for studies that can be included in best-evidence syntheses, with particular application to evaluations of middle and high school reading programs. These criteria include: comparison of children in classrooms using a given program to those in control classrooms using an alternative program or standard practice, studies using random assignment or matching with appropriate adjustments for any pretest differences (e.g., covariates), pretest data on the students and acceptable levels of difference in achievement between treatment and control classrooms (i.e., no more than 50% of a standard deviation), dependent measures that include a quantitative measure of reading performance, study duration of a minimum of 12 weeks, and at least two teachers and 15 students in each treatment group. This study meets each of these criteria, as outlined in the relevant paragraphs that follow.

The study was guided by three specific research questions:

1. What is the impact of an academic vocabulary program on the vocabulary and reading comprehension of language minority learners and their native English speaker classmates enrolled in urban middle schools?
2. With what level of fidelity was the program implemented and what do teachers report about ease of implementation?
3. In what ways did the instruction as implemented contrast with standard practice?

## Method

### Participants

The participants included 476 sixth-grade students (346 language minority learners and 130 native English speakers) and their teachers ( $n = 19$ ), from seven middle schools in an urban district in the southwestern United States. The student sample was 53% female, and the median age of participants at the beginning of the intervention period was 11 years and 11 months. At pretest, the mean score for the sample on the Gates–MacGinitie Reading Comprehension Test, Fourth Edition (2000) was the 35th percentile. The study design and selection procedure (described in the next section) yielded 12 teachers teaching 13 sections with 296 students in the treatment group, and 7 teachers teaching 8 sections with 180 students in the control group. Of note,

the initial sample included 580 students, however, 94 students did not complete their academic year in the same classroom or school, in large part due to within-school mobility; there were no differences on any pretest achievement measures between those students who left the study and those who remained.

Table 1 presents descriptive statistics on the demographics of the sample. As shown, participating students were ethnically and linguistically diverse, with approximately three fourths of students reporting a language other than English spoken at home. Information on language minority learner status and race/ethnicity was obtained by a survey administered to all participating students. Students who reported speaking English exclusively at home were classified as native English speakers, whereas students reporting that a language other than English was spoken at home to any degree

**Table 1. Sample Demographics ( $N = 476$ )**

Demographics (proportion in each category)	Treatment ( $n = 296$ )	Control ( $n = 180$ )	Total ( $n = 476$ )
Gender			
Female	54.1%	50.0%	52.5%
Ethnicity			
African or African American	5.7%	10.6%	7.6%
Asian	5.7%	6.7%	6.1%
Caucasian	16.9%	13.9%	15.8%
Hispanic/Latino	50.3%	46.1%	48.7%
Pacific Islander	1.4%	1.7%	1.5%
Multiracial/Not reported	19.9%	21.1%	20.4%
Home language use			
Only English	24.1%	27.0%	25.1%
Mostly English	13.2%	16.3%	14.4%
English and another language equally	33.2%	28.1%	31.3%
Mostly another language	22.0%	20.8%	21.6%
Only another language	7.5%	7.9%	7.6%
Home language(s) listed			
English (exclusively)	24.1%	27.0%	25.1%
Spanish	59.8%	60.0%	59.9%
Vietnamese	1.7%	2.2%	1.9%
Lao	1.4%	2.2%	1.7%
Hmong	1.4%	1.7%	1.3%
Somali	1.4%	1.7%	1.5%
Pilipino/Tagalog	1.7%	3.9%	2.3%
Other	6.4%	8.3%	6.7%
Immigration background			
Born in the U.S.	86.7%	81.8%	84.9%
Born in Mexico	7.9%	12.5%	9.6%
Born in a country other than U.S. or Mexico	5.6%	5.7%	5.5%
Attended school in the U.S. since K or Grade 1	95.0%	90.8%	92.4%

were classified as language minority learners. This broad definition for language minority learner includes students whose families speak English predominately and those who speak English and another language in equal amounts, as well as those from homes in which another language predominates, consistent with the definition of this population offered by the National Literacy Panel on Language Minority Children and Youth (August & Shanahan, 2006). Students reported one or more racial/ethnicity group(s) with which they identify from six categories, chosen based on the school district's demographic reports: African American, Asian American, Caucasian/White, Hispanic/Latino, Pacific Islander, or Other. All language minority learners in the study had received all of their instruction in English.

The participating teachers varied with respect to years of teaching experience and level of higher education, but this variation was comparable across the treatment and control groups. Three teachers had between 1 and 2 years of teaching experience, 4 teachers had between 3 and 5 years of teaching experience, 7 teachers had between 5 and 16 years of teaching experience, and 5 teachers had between 16 and 30 years of teaching experience. Ten teachers reported their highest degree to be a bachelor's degree, while 9 teachers reported having a master's degree. Thus, the median teacher for the sample had taught for between 5 and 16 years and held a bachelor's degree.

The seven participating schools were largely representative of schools in urban districts in California—districts in large cities, characterized by racial, linguistic, and economic diversity. The student population in each school comprised a large percentage of students of color (median = 66.9%; ranging from 43.5% to 96.4%) and of students from low-income backgrounds (median = 58.4%; ranging from 24.4% to 100.0%).

## **Study Design**

The study is quasi-experimental in nature. Having selected schools based on student population and the match between the program and the needs of a sizeable group of their sixth-grade students, the principals from participating schools opted into the study. Together with the principal, in each of the seven schools, we identified all mainstream sixth-grade English language arts sections (i.e., groups of students). We excluded sections that were specially designated as gifted and talented education sections or as structured English immersion sections for recent immigrants with very limited English proficiency. Then, we were able to examine the characteristics of the sections (e.g., student achievement, student demographics) to determine what would create an approximate match between the treatment and control conditions within each school. The principal then sought approval from this pool of teachers

(approximately three per school), with one or two teachers agreeing to deliver the academic vocabulary program intervention, and another teacher agreeing to have her or his students serve as the control group.<sup>1</sup> Because each sixth-grade teacher typically taught several sections of students, we sampled one section for each teacher. The sampling from each teacher's possible sections was random, though not independent for each teacher within a school; in each school, the section for the first teacher was selected at random and then the section for each remaining teacher was chosen at random from only those sections taught at a different time than the first section to facilitate efficient and cost-effective data collection. In the case of two teachers, we sampled two sections to capture a more representative range of their students' performance. Each of these two teachers had one section in which students with quite low achievement predominated as well as additional sections that were more representative of the overall population of the school. The number of available teachers and sections differed by school due to differing enrollment and numbers of excluded sections, yielding between two and four sections in each school.

## **Intervention Program**

The intervention was a text-based academic language program, which we refer to as Academic Language Instruction for All Students (ALIAS) developed for use in mainstream, low-performing English language arts classrooms with high numbers of language minority learners. The program was 18 weeks in length, featuring 8 two-week units, each consisting of an 8-day lesson cycle, and 2 one-week review units. Each daily lesson in the cycle was designed to be 45 minutes, with lessons delivered 4 days per week. These 45 minute lessons were implemented in the context of the participating schools' English language arts block, which lasted between 90 and 120 minutes a day. It is worth noting that at the time of the study, district curriculum leaders felt that 45 minutes was a substantial but appropriate investment of time, given their concern that teachers have trouble planning effectively for such long blocks and have received generally limited guidance on how to do so. As described in what follows and in the Appendix, instructional activities in the 8-day lesson cycle were designed to build knowledge of the words incrementally over time by providing multiple exposures to the words in different forms and in different meaningful contexts.

Each unit revolved around a short piece of engaging informational text—a feature article from *Time for Kids* magazine, to which the participating school district subscribes. We selected texts on the basis of several criteria: potential for student engagement, readability at the fourth- to sixth-grade instructional level, length,

and the opportunities available for teaching academic vocabulary. Several of the texts featured topics salient to adolescent youth culture, such as single-gender classrooms and television viewing rates, while others addressed issues of diversity, such as how different ethnic groups in Africa learn to get along.

As shown in Table 2, from each text, we chose 8 or 9 high-utility academic words that appeared in both the text and on the academic word list (AWL; Coxhead, 2000). The AWL is an empirically based compilation of words (e.g., *evidence, method, integrate*) that appear frequently in texts across many academic disciplines (distinct from discipline-specific words such as *paralelogram, antebellum, metaphor*) but occur more rarely in oral conversation and narrative texts; as such, the list represents words that are most worth teaching and learning across academic disciplines. The AWL was compiled from a corpus of 3.5 million words of written academic text by examining the range and frequency of words outside the first 2,000 most frequently occurring words of English. The AWL contains 570 word families that account for approximately 10% of the total words in academic texts but only 1.4% of the total words in a fiction collection of the same size. This academic domain of vocabulary thus represents high-utility words that appear commonly in expository text

but are not specific to any particular academic discipline. Originally generated based on an analysis of university textbooks (Coxhead, 2000), these words also appear frequently in sixth-grade content area textbooks (Nair, 2007). As noted, each unit focused on 8 or 9 academic words throughout the 8 days; a total of 72 words were taught in the program. Of note, to ensure the instruction was authentic, meaningful, and engaging, not every target word appears in every day's lesson; rather, selected words were used to process content and meet lesson objectives. Each unit provided between 3 and 4 exposures to each word and a range across each day of 2 to 5 times. Across the program, 11 words were target words in 2 units and thus students had more exposures to each of these words.

The 8-day cycle included a variety of whole-group, small-group, and independent activities designed to promote deep processing through opportunities for listening, speaking, reading, and writing with the words. Table 3 provides a description of each day in the cycle, including sample lesson components, the rationale as we provided it to teachers (which was listed in the materials under the heading What's the Point?), the rationale from research and scholarship for this type of activity, and brief sample excerpts from the teacher materials. In the process of implementing the general principles

**Table 2. Targeted Words (Repeated Words in Italics)**

Unit 1: Players With Pride	Unit 2: High-Tech Bullies	Unit 3: New Clues to a Mystery	Unit 4: Separated at School
Affect	Communicate	Ancient	<i>Community</i>
Culture	Identity / Identify	Area	Discrimination
Community	Incidents	Complex	Distinctions
Contribute	Legally	Integrated	Evidence
Establish	Method	Located	Gender
Ethnic/ethnicity	Policy	Major	Options
Residents	Research	Period	Regulations
Welfare	Require	Puzzle	Research
	Survey	<i>Researcher</i>	Respond
		Site	Since
		Which	<i>Survey</i>
			Topic
Unit 5: Bad News for Bees	Unit 6: Do Kids Tune in Too Much	Unit 7: Witness to History	Unit 8: Going the Distance
Aware	According	<i>Area</i>	Anticipate
Collapse	Average	<i>Awareness/aware</i>	Constantly
Conduct	Expert	Civil	<i>Contribute</i>
<i>Contribute</i>	Foundation	Documentary	Convince
Crucial	Media	Image	Effect
<i>Identify</i>	Nearly	Inspire	Expanse
<i>Research</i>	Percent (%)	Issue	Generate
Resource	<i>Survey</i>	<i>Research</i>	<i>Inspire</i>
Seeking		Social	<i>Image</i>
Theory		Survive	<i>Researcher</i>
Transport		Vision	Releasing
Widespread		While	Region
Yet			Survive
			Until

**Table 3. Lesson Cycle Description, Rationale, and Sample Instruction**

	Sample lesson components	Rationale offered to teachers (What's the point?)	Rationale from research and scholarship	Sample excerpts from teachers' materials
<b>Day 1</b>	<ul style="list-style-type: none"> <li>• Read article and discuss concepts.</li> <li>• Introduce target words.</li> </ul>	<p>Rather than simply memorizing vocabulary, students will be invested and engaged in learning concepts and acquiring background knowledge as they read the text, and then they will begin to map the labels (words) onto these new concepts. Using the text to introduce an engaging topic also provides a meaningful context for discussing the words throughout the unit.</p>	<p>A range of studies have supported the importance of both definitional and contextual information (Stahl &amp; Fairbanks, 1986; Stahl &amp; Nagy, 2006).</p>	<p>Students discuss</p> <ul style="list-style-type: none"> <li>• What have teachers in the article noticed happening in the single-gender classrooms?</li> <li>• Do teachers in Texas believe that this experiment is working?</li> <li>• Do you think you would like this class to be a single-gender classroom?</li> </ul>
<b>Day 2</b>	<ul style="list-style-type: none"> <li>• Introduce how to use context clues as a word-learning strategy.</li> <li>• Brainstorm word meanings.</li> <li>• Create class definitions using accurate information.</li> <li>• Create and record personal definitions by rephrasing class definition in students' own words.</li> </ul>	<p>With input and discussion about the words' meanings from the text, the teacher and their classmates, students will begin to form an accurate understanding of the words. Teacher will be able to build on students' prior knowledge while immediately clarifying any misinformation as the students share their ideas aloud.</p>	<ol style="list-style-type: none"> <li>1. Cognitive research indicates the importance of schema activation prior to the development of more complex schema (Bransford, 2004).</li> <li>2. Practice-based scholarship recommends activating and building on students' word knowledge prior to introducing formal definitions, in order to refine vague understandings and correct misunderstanding (Marzano &amp; Pickering, 2005).</li> <li>3. "Definitions can be confusing to children... Having children restate definitions may be the only way a teacher can find out whether the children actually understand them." (Stahl &amp; Nagy, 2006, p.65).</li> </ol>	<p>Context clues reviewed: *Let's do some more work with context clues. Look at these two sentences using a familiar word, scores, and let's practice how to use context to determine meaning:</p> <ul style="list-style-type: none"> <li>• The new player on the soccer team scores a goal every game.</li> <li>• Sonia has the best test scores in the class.</li> </ul> <p>Informal Definitions:</p> <ul style="list-style-type: none"> <li>• On scrap paper, students write down all they know about each word.</li> <li>• Students share word knowledge; the teacher confirms and posts accurate information and contributes necessary additional information to define each word.</li> <li>• Class comes up with informal definitions using information generated together.</li> </ul>
<b>Day 3</b>	<ul style="list-style-type: none"> <li>• Answer text-based questions using target words as a whole-group and in pairs.</li> <li>• Share answers with classmates orally.</li> </ul>	<p>Students apply information drawn from dictionary and personal definitions as they work together to answer literal and inferential questions from the text. This allows them to use target words in a scaffolded, highly contextualized situation, first orally, and then in writing.</p>	<ol style="list-style-type: none"> <li>1. Several studies have highlighted the importance of deep processing through oral and written activities (Beck et al., 2002; Carlo et al., 2004; Graves, 2006; Stahl &amp; Nagy, 2006).</li> <li>2. "When you want to know about students' ability to use a new term correctly... you can ask students to use vocabulary in meaningful ways in the context of some larger activities. The most direct way to do this is to ask students to incorporate particular words in their responses to questions, and their summaries and retellings... evaluating their vocabulary usage in the most authentic way." (Blachowicz &amp; Fisher, 2007, p. 193).</li> </ol>	<p>Directions to students, plus sample Reader Response questions: With a partner, discuss the following questions before writing your own answer in your notebook:</p> <ul style="list-style-type: none"> <li>• What <i>evidence</i> will the school officials use to decide if the program is working and should be continued?</li> <li>• How do students <i>respond</i> differently in single-gender classrooms?</li> <li>• What kind of <i>discrimination</i> against girls happened regularly before Title IX?</li> </ul>

(continued)

**Table 3. Lesson Cycle Description, Rationale, and Sample Instruction (continued)**

	Sample lesson components	Rationale offered to teachers (What's the point?)	Rationale from research and scholarship	Sample excerpts from teachers' materials
<b>Day 4</b>	<ul style="list-style-type: none"> <li>• Sketch representation of target word and write related sentence.</li> <li>• Complete target word practice task to increase exposure to new vocabulary in different contexts (crossword puzzle, sentence judgment tasks).</li> </ul>	<p>This task provides a different modality for thinking about word meaning. Students move toward a much deeper understanding of the word by making a productive representation of each target word and creating a sentence that uses the word accurately. Both of these tasks entail a metacognitive aspect, namely “knowing that I know enough to represent this word in a picture...knowing that I know enough to recognize when the word makes sense in a sentence.” On this day, students work on an independent task, which allows them to focus on their own word knowledge, and also allows the teacher to assess understanding.</p>	<ol style="list-style-type: none"> <li>1. Creating scenarios around words as contexts for deep processing is recommended (Stahl &amp; Nagy, 2006).</li> <li>2. “When you ask students to construct a picture, symbol or graphic representation of a term, they are forced to think of the term in a totally different way” (Marzano &amp; Pickering, 2005, p 21).</li> <li>3. Asking students to make judgments about the sensical and nonsensical use of words in sentences provides an opportunity for deep processing (Carlo et al., 2004).</li> </ol>	<ul style="list-style-type: none"> <li>• Teacher thinks aloud while sketching the word <i>discrimination</i>: “I know that <i>discrimination</i> refers to when a person is treated unfairly for any number of reasons, such as age or gender or race or religion. This is a bit of a hard one to draw, but I think I am going to make a picture of a kids’ clubhouse and I am going to put a sign on the door that says, “No girls allowed!” While this is just a kids’ clubhouse, it is still <i>discrimination</i> described here because the girls are not allowed in just because they are girls.”</li> </ul>
<b>Day 5</b>	<p>Morphology lesson:</p> <ul style="list-style-type: none"> <li>• Teach specific target suffixes directly.</li> <li>• Lead whole-class discussion to complete a word form chart using forms of the target words.</li> <li>• Provide student practice with making and recognizing word forms in writing and revising activities related to the theme of the text.</li> </ul>	<p>The purpose of this lesson is to provide direct instruction in word forms. Teaching morphology gives students an understanding of how words are related and how they can figure out unknown words by using word parts. This lesson addresses both word-specific knowledge (e.g., different forms of the target words) and word-general skills (e.g., how to be metacognitive about breaking down words).</p>	<p>Studies have shown the benefits of morphology instruction (Baumann et al., 2002; Carlo et al., 2004), leading scholars to recommend this strategy (Stahl &amp; Nagy, 2006).</p>	<p>Direct teaching of a suffix: <i>-al</i> is like the other suffixes we’ve learned because it changes the part of speech of the word. <i>-al</i> changes words from nouns (person, place, thing, or idea) into adjectives (describing word). So, if there is a person who does magic, I would describe her by saying that she is <i>magical</i>. <i>Magic</i> is a thing or a noun, but <i>magical</i> is the way you would describe someone or something, so that makes <i>magical</i> an adjective.</p> <p>Excerpt from Find the Misfits practice activity: [Students read paragraph and correct incorrect word forms] As a student at an all-girls middle school since sixth grade, I believe that there are both advantages and disadvantages to this <i>typical</i> of school. The advantages are that teachers can make sure we learn with the <i>methodical</i> that are best for us....</p>
<b>Day 6</b>	<ul style="list-style-type: none"> <li>• Use the words in new contexts by answering questions based on new scenarios (e.g., pairs of students complete a mock interview by answering questions that use the target words, from the point of view of a character they’ve selected.</li> <li>• Teach multiple meanings of target words.</li> </ul>	<p>Students need to use the target words outside of the context of the article to deepen their understanding of the words. As they answer questions as a famous person during a mock interview, they will use the target words in new contexts, and their knowledge of the words will be refined. Students need opportunities to distinguish between different meanings of words and different shades of a word’s meaning.</p>	<p>Scholars highlight that negotiating multiple meanings and multiple shades of meaning for a word is important to successful recognition and use of word meanings (e.g., Stahl &amp; Nagy, 2006).</p>	<p>Mock interview sample questions to be answered in character:</p> <ul style="list-style-type: none"> <li>• Are you an <i>expert</i> on any subject? Tell me about a topic you know a lot about.</li> <li>• When people see you in the <i>media</i>, do you think that they get information that describes you as you really are? Why or why not?</li> <li>• The Foundation for Homeless Children and the Foundation to Prevent Global Warming both want you to help them raise money. Which Foundation would you rather help? Why?</li> </ul>

(continued)

**Table 3. Lesson Cycle Description, Rationale, and Sample Instruction (continued)**

	Sample lesson components	Rationale offered to teachers (What's the point?)	Rationale from research and scholarship	Sample excerpts from teachers' materials
<b>Day 7</b>	<p>Prewriting Work:</p> <ul style="list-style-type: none"> <li>Teachers review the prewriting process, modeling a sample response using student input.</li> <li>Students talk through their ideas with a peer and use graphic organizers to categorize their ideas.</li> </ul>	<p>Students will engage in a highly supported process for planning and organizing their writing, in order to help them create a coherent piece of persuasive writing that uses some of the target words.</p>	<p>Skilled writers are more knowledgeable about the process of writing than less skilled writers (e.g., Saddler &amp; Graham, 2007).</p>	<p>Example of a writing prompt—agree or disagree? Respond to one of the following statements:</p> <ul style="list-style-type: none"> <li>Parents who want single-gender education for their children should send them to private schools.</li> <li>All middle schools should have single-gender classes because then all students would be successful.</li> <li>Title IX has made all school sports programs fair and equal for all students.</li> <li>Students who attend single-gender schools will be more successful in life.</li> </ul>
<b>Day 8</b>	<ul style="list-style-type: none"> <li>Write paragraph, revise &amp; edit (Teacher models moving from a graphic organizer to a extended piece of writing. Students create a paragraph based on their prewriting work, and then complete a self-assessment checklist to help them think through the steps in the revision process.)</li> <li>Share paragraphs with peers.</li> </ul>	<p>On this final day in the unit, students will use the target words in complete sentences within a coherent piece of writing. To do this, students must completely understand the concept behind the word's label, as well as how the word fits properly into the structure of various sentences. When students can successfully incorporate target words into their writing, teachers can be assured that the students know the words deeply.</p>	<p>Given the reciprocal relationship between reading and writing (e.g., Graves &amp; Watts-Taffe, 2002), authentic opportunities to use words in extended writing can serve as an opportunity for consolidating knowledge of word meanings.</p>	<p>Other sample writing prompts (target words in italics):</p> <ul style="list-style-type: none"> <li>What <i>crucial</i> invention from the past one hundred years has most <i>affected</i> your daily life?</li> <li>What strategies do kids use to <i>survive</i> middle school?</li> <li>In your opinion, is a bike law a necessary <i>regulation</i>?</li> <li>Pretend you are going to have a family reunion and you can all meet anywhere in the world. Where would the <i>site</i> of your family reunion be?</li> </ul>
<b>Review Units</b>	<ul style="list-style-type: none"> <li>Engage in game-like activities to review the target words in all previous units.</li> <li>Reteach specific words based on student assessments.</li> </ul>	<p>Students need multiple opportunities to expand and consolidate word knowledge by applying vocabulary in new contexts.</p>	<ol style="list-style-type: none"> <li>"The aim of rich instruction [is] to have students engage in active thinking about word meanings, about how they might use the words in different situations, and about the relationships among words." (McKeown &amp; Beck, 2004, p. 18).</li> <li>Given the incremental nature of vocabulary learning (Stahl &amp; Nagy, 2006), opportunities for reteaching specific words are valuable.</li> </ol>	<p>Example of a review unit task, Interactive Crossword:</p> <p>Students work in pairs to complete a crossword; one student has all the across words filled in, the other has only down words filled in. Students take turns defining each target word for their partners who then use the word to complete items on their crossword.</p>

for effective vocabulary instruction described above, we drew on reports from individual experimental and quasi-experimental studies (e.g., Baumann et al., 2003; Carlo et al., 2004), but also from the fuller descriptions of activities available in research-based books for teachers (e.g., Stahl & Nagy, 2006). In a few cases to select

additional activities for deep processing of academic words, we drew on promising approaches recommended by practice-based scholarship (e.g., Blachowicz & Fisher, 2007; Marzano & Pickering, 2005).

As can be seen by examining the lesson components for each day, the instructional cycle follows a

developmental sequence of specific activities for building up word knowledge incrementally, beginning with exposure to the word through text, activation of students' prior knowledge of the word, work on the target word's meaning in the context of the text, introduction of additional meanings of the word, a morphological analysis of different forms of the word, and finally, use of the word in students' own writing. Figure 1 provides a visual metaphor of a brick wall that was used to illustrate for the teachers how the lesson cycle built vocabulary knowledge incrementally over the eight days. The one-week review units included cooperative games focused on the previously taught words as well as opportunities to reteach specific words or morphological skills with which students were struggling, identified based on assessments used during the week. The first review unit featured words from the first of half the program and the second featured words from across the entire program. The program materials provided to teachers included a map of the eight-day cycle, a quick reference sheet for the activities, a one-page outline for each daily lesson listing the essential components, and a more elaborated instructional model or sample script for potential teacher and student talk for each activity; the one-page outline and the instructional model are provided as examples in the Appendix.

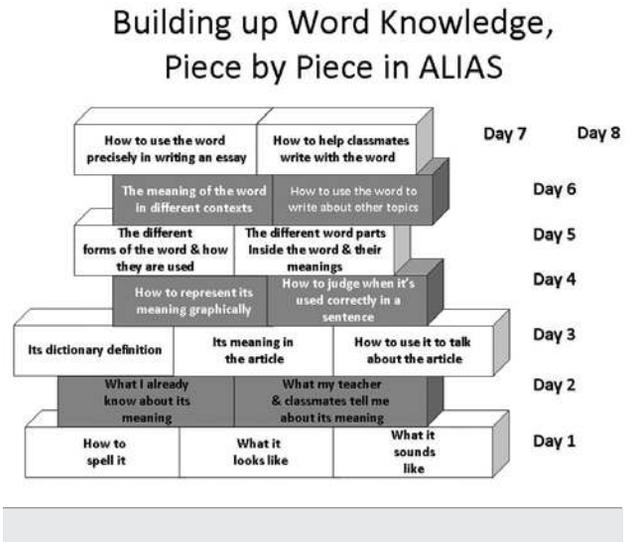
To support the implementation of the intervention, teachers met with a program specialist—a former bilingual teacher with extensive experience in the district who had been trained by the research team—on a monthly basis. The meetings were designed to be supportive and based on teachers' implementation needs, with an organic format in which the program specialist addressed questions and issues, from minor logistical details to more substantive issues about the program, such as troubleshooting difficult lessons. In some cases, these meetings were conducted with more than one teacher at the same school and thus also provided an opportunity for teachers to discuss the instruction with their colleagues.

## Measures

### *Student Measures of Language and Reading*

We administered a battery of standardized and researcher-created assessments to students at pretest and posttest. The batteries included the same measures with the exceptions of the Word Association and Word-Meanings-in-Context tasks, which were administered at posttest only due to time constraints. For standardized, norm-referenced assessments, we used equated, distinct forms at the two time points to minimize

Figure 1. The ALIAS Instructional Cycle



practice effects. For researcher-created assessments, we used forms with the same test items (rearranged in different orders to minimize their surface similarity); though this may introduce practice effects, these effects can be assumed to be the same across treatment and control conditions. The researcher-created assessments were piloted prior to pretest with 93 students in four sixth-grade classrooms in two nonparticipating schools that were comparable in demographics and achievement levels to participating schools; as described below, the final versions of these measures included selected items with the greatest evidence of reliability (i.e., internal consistency within measure) and validity (i.e., correlations with standardized measures of vocabulary and reading comprehension) from the pilot test. As a result of this process, the final measures included 53 (or 74%) of the 72 target words taught in the program.

### **Vocabulary**

Reading vocabulary was assessed using one standardized, norm-referenced measure and four researcher-created instruments.

#### **Stanford Achievement Test-10th Edition: Reading Vocabulary Subtest**

(The Intermediate II, fall of sixth grade form was used at pretest and the Intermediate III, spring of sixth grade form was used at posttest.) This standardized, norm-referenced test is a widely used assessment of grade-level reading vocabulary knowledge in which students are asked to identify synonyms for a given word, identify the meaning of a rare word in a sentence context, and match sentences in which a word with multiple meanings is used in the same way. The publisher provides

evidence of good reliability (Kuder–Richardson Formula 20 reliability = 0.86; alternate forms  $r = 0.77$ ) and extensive evidence of validity.

### Target Word Mastery

This researcher-created measure is a 32-item multiple-choice task in which students choose a synonym for a given word drawn from the words taught in the curriculum. During pilot testing, 45 items were created based on 45 words randomly selected from the total 72 target words, and the 32 items with the most evidence of reliability and validity were selected. The estimate of internal consistency reliability for this task in the current sample prior to treatment was good (Cronbach's  $\alpha = 0.85$ ). A prior study using a version of this task provided convergent and divergent validity evidence and extensive research provides evidence for the validity of this commonly used method for assessing vocabulary.

### Word Association

Depth of vocabulary knowledge of the target words was assessed with the Word Association task. Drawing on the design of tasks used in prior research (Carlo et al., 2004; Schoonen & Verhallen, 1998, cited in Carlo et al., 2004), this task taps students' knowledge of paradigmatic associations (e.g., that *effect* can be substituted for *consequence* while preserving meaning) and syntagmatic associations (e.g., that an *effect* can be *caused*). The task consisted of 15 items, each of which drew on knowledge of words taught in the program. During pilot testing, 30 items were created based on 30 target words randomly selected from the 50 target words not assessed on the Word-Meanings-in-Context or Morphological Decomposition tasks; based on pilot results, 15 items that had appropriate distracters and that reliably tapped this single construct were selected. Each target word appeared in the center of a box, surrounded by six other words, three of which were immutably associated with the target word, and three of which were only circumstantially related to the target word. For instance, *effect* has immutable associations with *cause*, *consequence*, and *result*, but only circumstantial associations with *negative*, *policy*, and *people*. Students were instructed to choose the three words that “always go with the target word” and provided with feedback on two practice items using common words (i.e., *foot* and *dog*). The estimate of internal consistency for this task with the current sample prior to treatment was adequate<sup>2</sup> ( $\alpha = 0.78$ ). In addition to validity evidence provided by prior studies (Carlo et al., 2004; Schoonen & Verhallen, 1998, cited in Carlo et al., 2004), convergent evidence from the current sample prior to treatment includes strong correlations with the other standardized and researcher-created measures of vocabulary knowledge ( $r = 0.67$  to  $0.69$ ) and divergent

evidence includes weaker correlations with a nonword morphological awareness measure ( $r = 0.47$ ).

### Morphological Awareness

Students' morphological skills were assessed using the Morphological Decomposition Task, an instrument that we created based on previous research (Carlisle, 2000; Carlo et al., 2004). In this task, testers provide students with a word with a derivational suffix (e.g., *complexity*) and ask them to extract the base word (e.g., *complex*) to complete a sentence (e.g., The problem is \_\_\_\_\_). The task included 18 items, 9 of which included words taught in the program. The nine target words were selected from those not assessed on the Word-Meanings-in-Context measure that could be derived clearly with the three suffixes chosen (*-ity*, *-sion/-tion*, *-al*), which were judged to be the three most challenging and useful suffixes taught in the program. Test administrators read the word and sentence aloud to minimize the influence of word reading skills, and students responded with a written answer. Trained research assistants scored written answers to the task dichotomously using a detailed scoring guide that included a rubric along with sample correct and incorrect responses. Responses were scored as correct if they provided the correctly spelled form of the word or a phonetically justifiable version of the word form, such as *possess* for *possess* or *durible* for *durable*. Responses were scored as incorrect if they were morphologically unrelated words such as *have* for *possess* or *hard* for *durable*, when they were incorrectly decomposed responses such as *poss* or *dura*, or when they were ambiguous responses such as *possese* and *durabil*. In this way, we protected in part against the confounding of variation in students' ability to spell the base word with true variation in morphological awareness. The estimate of internal consistency reliability at posttest was appropriately high (Cronbach's  $\alpha = 0.82$ ), and interrater reliability, calculated based on double coding of 20% of the posttests, was very high (Agreement = 98%; Cohen's Kappa = 0.96). Several prior studies provide validity evidence for similar tasks using this item format, including convergent and divergent validity evidence (Carlisle, 2000; Carlo et al., 2004) and evidence of construct validity based on confirmatory factor analysis models (Wagner, Muse, & Tannenbaum, 2007).

### Word-Meanings-in-Context

Comprehension of target word meanings in the context of extended expository texts was assessed using the Words-Meanings-in-Context task, which was designed specifically for this study drawing on the framework proposed by Pearson, Hiebert, & Kamil (2007). To complement the more decontextualized vocabulary measures, this measure sought to assess students'

knowledge of the meanings of words in specific contexts as well as their global comprehension of texts that included the words taught. This task included both questions about target word meanings in context and questions about the meaning of the overall text; as such, it can be considered a contextualized vocabulary measure with a strong reading comprehension component. Although we had initially considered separating scores from the word-specific and text comprehension items, principal components analysis indicated that these two types of items tapped the same, unitary construct, and that such subscores would not yield reliably different information (details are available from the authors).

This task included five expository passages drawn from *Time for Kids* magazine that were candidates for inclusion in the program, but that do not appear in the final version. Students read each passage independently and answered multiple-choice comprehension questions during a 45 minute period, which was sufficient for all but a few students to complete the test. The items for each passage included one item measuring global comprehension of the passage (e.g., What is the main idea of this text?), one item that required students to make an inference across several statements in the passage (e.g., What do Alex, Joshua, and Nathan have in common?), and three items that tapped understanding of a taught target word in the context of the passage (e.g., What does *major* mean in the text? from a passage describing a teenager who can discuss every major Presidential candidate from 1896 to 2004). For the latter item type, the correct choice was an appropriate synonym (e.g., *important* for *major* in this context), whereas the distractors included another meaning for the target word that is not the meaning used in the passage (e.g., *military officer*), a word that is related to the content of the passage, but is not a meaning of the target word (e.g., *Republican*), and a word that looks similar to the target word and has a loose semantic relationship to the content of the passage (e.g., *majority*). As originally written, each passage contained three taught target words, which were thereby selected for assessment. After excluding two items that had virtually no correlation with the other items, the estimate of internal consistency reliability for the resulting 23-item task at posttest was adequate ( $\alpha = 0.80$ ). As convergent validity evidence in the current sample at posttest, the task had the strongest correlation with the standardized measure of reading comprehension ( $r = 0.70$ ) and appropriately strong correlations with the vocabulary measures ( $r = 0.57$  to  $0.67$ ). As divergent validity evidence, the lowest correlation for the task was with a morphological awareness measure that included nonwords ( $r = 0.45$ ; details on this task are available from the authors).

## **Reading Comprehension**

### **Gates–MacGinitie Reading Test, Fourth Edition: Reading Comprehension**

This standardized, norm-referenced measure is a widely used assessment of students' global reading comprehension, in which students have 35 minutes to read several grade-level passages from expository and narrative texts and complete multiple-choice questions. We administered the sixth-grade version published in 2000. Form S was used at pretest and form T was used at posttest. The publisher reports Kuder–Richardson Formula 20 reliability coefficients of 0.90 to 0.92 for the sixth-grade test, as well as extensive validity evidence.

### **Classroom-Level Measures: Fidelity of Implementation**

Fidelity of implementation was estimated and assessed using three methods: a weekly log completed by every teacher in the treatment group, classroom observations conducted in the treatment classrooms prior to and during the implementation of ALIAS, and classroom observations conducted in the control classrooms for the duration of the study.

### **Implementation Logs**

Treatment teachers completed weekly implementation logs ( $n = 18$  per teacher, total = 216 logs) through a password-protected website designed by the research team. On each weekly log, teachers completed four sections, one for each of the four daily lessons for that week. Each daily lesson section contained five subsections with the following headers:

1. Date of delivery
2. Minutes spent on ALIAS
3. Activities completed with a checklist of the three to six lesson components (e.g., preview and read aloud article, introduce target words)
4. Notes on strengths (optional; e.g., activities that were very effective or engaging)
5. Notes on weaknesses (e.g., activities that took longer than expected or did not meet the objectives)

Complete data was available for 95% of the logs, while the remaining 5% were missing due to a teachers' maternity leave of absence. The majority of logs (84%) were completed by the teachers on time on a weekly basis, while the remaining logs were completed retrospectively within approximately 4–6 weeks of the end of the unit, in response to individual requests from a member of the research team.

## Classroom Observations in Treatment Classrooms

Trained research assistants observed treatment teachers on five to seven occasions during the intervention period, using an observation protocol with which they rated teachers' implementation of the intervention for fidelity (i.e., the presence of each of the three to six lesson components for a given daily lesson) and quality (i.e., ratings of low, medium, or high on a range of nine instructional quality variables specific to the intervention). Observations occurred every two to three weeks during the implementation period. We scheduled observations systematically to capture the same day in the same unit (e.g., Day 5 in Unit 5) across all teachers and to capture seven of the eight different daily lessons in the ALIAS lesson cycle and seven of the eight different units. Ratings for fidelity and quality were averaged across items and across observations to yield composite scores for each teacher. Thus, the reliability of the composite scores depends on both consistency among raters and consistency among items and can be summarized with a single Cronbach's  $\alpha$  statistic. The estimate of reliability for the teacher-level composites were good for fidelity of implementation (0.76), quality of intervention-specific instruction (0.77), and quality of general instruction (0.85).

## Classroom Observations in Control Classrooms

Control teachers were observed once prior to the start of the intervention period and three to four times during the intervention period. These observations lasted for 45 minutes and were divided into three 15-minute intervals, during which observers took field notes to provide rich description of the instruction taking place on an observation protocol. An independent coder, who was a graduate research assistant with prior teaching experience, coded the curricular content of each 15-minute interval based on these field notes as well as on video recordings of the observation when necessary. The coder used an a priori coding scheme with 10 categories for curricular content (e.g., vocabulary, reading comprehension with expository text, writing conventions) developed based on the categories used in the California state curricular content standards for English language arts. To estimate interrater reliability, the second author independently coded 20% of the protocols and attained 100% agreement with the original codes. Observers also coded any incidental vocabulary instruction observed on the observation protocol, using a coding scheme adapted from previous research (Gersten, Dimino, & Jayanthi, 2007).

After the completion of the 45-minute observation, observers rated the control teachers on 15 instructional

quality indicators that were equivalent to those used in the treatment classrooms. Of these, 9 indicators were specific to the instructional approach used in intervention, although potentially observable in any classroom (e.g., Teacher affirms correct word definitions and usages; Teacher facilitates student talk) and 6 indicators were related to general instructional quality (e.g., Teacher is prepared for class; Teacher responds effectively to misbehavior). Ratings of instructional quality were averaged across items and observations, yielding two composites for each control teacher. Each control teacher was observed by two different observers, thus the reliability of the composites depends on both interrater and interitem consistency, which was estimated to be high (intervention-specific instructional quality, Cronbach's  $\alpha = 0.84$ ; general instructional quality, Cronbach's  $\alpha = 0.84$ ). The raters were not blind to the condition of the class they were observing, due to logistical concerns in observing the treatment teachers on specific days of the program and the necessity of ensuring that these observers were sufficiently familiar with the program to rate fidelity, as well as financial constraints in staffing the project efficiently.

## Teacher Survey

To gain insight into how the ALIAS program may have influenced instructional practices, at the end of the study participating teachers completed a survey, in which they reported retrospectively on the frequency of specific classroom activities. The survey asked teachers to think back to a typical month of instruction and to report on how frequently the type of activity or instruction occurred in their classroom on a scale that included never, 1–2 times per week, 2–3 times per week, and 3–4 times per week. The items included 34 specific curricular elements (e.g., oral reading accuracy, word choice in writing) and 32 instructional methods (e.g., using graphic organizers, reading aloud) and types of student participation (e.g., discussing texts in pairs, completing a language arts worksheet). Treatment teachers reported both the frequency with which they typically included these elements in their instruction and on whether they spent more or less time on them during the treatment period. Eleven out of 12 treatment teachers and 6 out of the 7 control teachers provided complete data, while the remaining 2 teachers were unavailable at posttest due to medical leaves.

## Teacher Interview

During the posttest, every treatment teacher was interviewed for approximately 30 minutes. Semistructured interviews were conducted one-on-one by two members of the investigator team using a common protocol, which was developed through an iterative process of

drafting, piloting with nonparticipating teachers, and revision. The protocol included a presentation of the purpose of the interview, an assurance of confidentiality, and questions grouped around three broad categories derived from our research questions:

1. How was the program different from typical instruction?
2. What were barriers and facilitators of continuing use of the program?
3. What support for implementation was helpful or needed?

Specific questions within these categories explored the teachers' perceptions and use of the program in relation to its appropriateness given their students' needs, student response to instructional activities, and the specific supports that had been available to participating teachers. In addition to open-ended questions under these categories (e.g., *What support did you find helpful this year?*), the protocol included related probes to be used to follow up when teachers did not volunteer responses on specific targeted issues (e.g., *Did you collaborate with other teachers at your school who were teaching ALIAS?*). Following semistructured procedures, interviewers also asked participants to elaborate on and clarify comments to explore how the participants made meaning of their experiences. All interviews were audio recorded with the teachers' permission and subsequently transcribed.

### **Procedure**

The study was conducted over the course of the 2007–2008 academic year. The classroom observations began in the fall and the pretest was conducted in December. The posttest took place within two weeks of completing the ALIAS program and was conducted between the end of May and July; five of the seven schools were on a traditional calendar and thus students took the posttest between the end of May and mid-June, while the other two schools were on a year-round schedule and thus students took the posttest in mid-July. Trained graduate research assistants with teaching backgrounds administered the student assessments, which were group-administered in every participating classroom. The interviews with treatment teachers were conducted in a quiet room outside of the classroom, while the teacher's students were taking the posttest assessments.

### **Analytic Plan**

Quantitative and qualitative analytic techniques were employed to address each of the three research questions. First, descriptive statistics were calculated to determine means and standard deviations for the student assessment data, followed by a series of multilevel

models (also known as hierarchical linear models), which were fitted to investigate the impact of the intervention on student performance. Next, quantitative and qualitative analyses were undertaken to address the question about ease of program implementation. Descriptive statistics were calculated to investigate fidelity of implementation, including minutes of instruction and ratings of the quality of implementation and instruction. In turn, the results of the teacher survey were tabulated and qualitative coding was undertaken with the 593 total comments in the weekly log entries and the 11 teacher interviews conducted.<sup>3</sup>

The qualitative data were designed to complement the quantitative data by allowing for a process analysis to better understand the implementation of the intervention and its mechanisms (Yoshikawa, Weisner, Kalil, & Way, 2008). This is particularly important given the situated complexities of teaching and learning (Sumara, 2000). To analyze the interview data, we largely relied on an open coding approach, stemming from the grounded theory method (Maxwell, 2005; Strauss & Corbin, 1998). A close initial reading of the transcripts was used to identify emergent themes that cut across the interviews as teachers described the implementation process. These emergent themes, together with the a priori themes reflected in the questions, formed the basis for our coding. A similar process was undertaken with the weekly logs which, because of their more open-ended structure, were coded primarily according to emergent themes. The coding was primarily focused on the comment responses provided in the columns labeled "Strengths" and "Weaknesses" for each lesson in every unit, and in the more general "Notes" section that appeared on every log. Throughout our analysis of both forms of qualitative data, we engaged in an iterative process that included building and revising theories and returning to the data to verify and refine our understanding of teachers' experiences of the implementation<sup>4</sup>.

---

## **Results**

### ***Impact on Student Achievement***

As a precursor to the analyses carried out to determine the impact of the instruction, multilevel models with random effects for classroom and school were fitted to investigate whether students in the treatment and control groups significantly differed at pretest on any of the measures. Likelihood ratio tests indicated that there were no statistically significant differences at pretest (SAT-10 Reading Vocabulary:  $p = .2674$ ; Target Word Mastery:  $p = .0811$ ; Gates–MacGinitie Reading Comprehension:  $p = .0947$ ; Morphological Decomposition:  $p = .5610$ ). The magnitude of the raw

differences between treatment and control indicated that the control group was (nonsignificantly) higher than the treatment group on each measure at pretest; these slight differences conform to the criteria for best evidence set forth by Slavin et al. (2008), described previously. The left three columns in Table 4 display the pretest scores on the student outcome measures for the entire sample and by treatment and control.

To evaluate whether the intervention had an impact on students' performance, we fitted a sequence of multilevel models in which the posttest score for each measure was regressed on a dummy variable representing condition (treatment or control). Specifically, we fitted three-level models to account for the nesting of children within classes as well as the nesting of classes within schools; this model has benefits over a simple ANCOVA model or a two-level model in that it produces standard errors and corresponding inference tests that are not biased by improper assumptions for the residuals given the hierarchical organization of the data. To improve the precision of the estimate of the treatment effects, we included several covariates. These included: pretest scores for each of the measures described above, centered at the student-level grand mean; a set of dummy variables to represent ethnicity, with the largest ethnicity, Latino/Hispanic, specified as the reference category; and a dummy variable for language minority

(LM) learner status, with native English speaker (NE) specified as the reference category.

For example, the hypothesized multilevel model for the Word Mastery outcome is given by the following equations. The level-1 (student-level) equation is:

$$POST\_TARGETWORDMASTERY_{ijk} = \beta_{0jk} + \lambda_1 Z_{ijk} + \epsilon_{ijk}$$

The level-2 (class-level) equation is:

$$\beta_{0jk} = \beta_{0k} + \beta_1 TREAT_{jk} + u_{jk}$$

The level-3 (school-level) equation is:

$$\beta_{0k} = \beta_0 + u_k$$

For simplicity, these three equations can then be collapsed into the following composite form:

$$POST\_TARGETWORDMASTERY_{ijk} = \beta_0 + \beta_1 TREAT_{jk} + \lambda_1 Z_{ijk} + (\epsilon_{ijk} + u_{jk} + u_k) \text{ where } \epsilon_{ijk} \sim N(0, \sigma_1^2), \text{ and } u_k \sim N(0, \sigma_2^2)$$

$POST\_TARGETWORDMASTERY_{ijk}$  represents the posttest score on Word Mastery for child  $i$  in class  $j$  in school  $k$ . Parameter  $\beta_0$  is the overall intercept. Parameter  $\beta_1$  represents the main effect of treatment on the posttest score for Word Mastery. Term  $\lambda_1$  represents a vector of parameters for the effects of the student-level control variables  $Z$ , listed above. Residual  $\epsilon_{ijk}$  represents the random effect for child  $i$  in class  $j$  in school  $k$ , residual  $u_{jk}$  represents the random intercept for class  $j$  in school

**Table 4. Results on Student Outcome Measures, expressed as Sample Means at Pretest, Overall and by Treatment and Control, and as Fitted Means at Posttest by Treatment and Control, Adjusted for Pretest Achievement and Demographic Covariates based on Multilevel Models ( $N = 476$ )**

Outcome measure	Pretest			Posttest		
	Overall sample mean	Control sample mean	Treatment sample mean	Control adjusted mean	Treatment adjusted mean	Treatment effect (Cohen's $d$ )
Target Word Mastery (raw score)	19.59	21.09	18.67	21.63	24.03	0.39***
Morphological Decomposition (raw score)	11.57	11.74	11.47	12.89	13.69	0.22***
Target Word Association (raw score)	33.41 <sup>a</sup>	34.97 <sup>a</sup>	32.65 <sup>a</sup>	35.87	36.89	0.15
Word-Meanings-in-Context (raw score)				12.94	13.88	0.20*
Gates-MacGinitie Reading Comprehension (Extended Scale Score; Spring of Sixth Grade Mean = 522)	499.79	506.00	496.01	498.50	503.32	0.15
SAT-10 Reading Vocabulary (developmental scale score)	634.26	639.26	631.23	643.82	644.01	0.005

<sup>a</sup>The pretest for the Target Word Association was conducted one month prior to the other pretests with a slightly different sub-sample of 405 participants, and was therefore not included as a covariate in estimating treatment effects. As a result, these estimates should be interpreted with caution.  
\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

$k$ , and residual  $u_k$  represents the random intercept for school  $k$ ; each of these residuals is independently drawn from a normal distribution with mean 0 and an unknown variances  $\sigma_\epsilon^2$ ,  $\sigma_1^2$ , and  $\sigma_2^2$ , respectively. Thus, the intercept was allowed to vary by class and by school, while all other effects were specified to be fixed across classrooms and schools. In particular, the effect of treatment was fixed to be the same across teachers and schools, consistent with our interest in the overall average effect of the treatment.<sup>5</sup> Class- and school-level variance components were found to be statically significant as demonstrated by Wald Chi-Square tests in both unconditional and final multilevel models. The effect of treatment and other effects described below were tested using likelihood ratio tests.

Results of multilevel modeling indicated that there were positive and statistically significant effects of the treatment on Target Word Mastery, Morphological Decomposition, and Word-Meanings-in-Context as well as positive and marginally significant effects of the treatment on Gates–MacGinitie Reading Comprehension and Target Word Association. Table 4 displays the adjusted posttest means for the treatment and control groups and the effect sizes for the treatment, and Table 5 presents the fitted multilevel models. As we would expect, the effect of treatments was largest for the most proximal vocabulary measure, Word Mastery ( $d = 0.39$ ;  $p < .0001$ ), and somewhat smaller for Morphological Decomposition ( $d = 0.22$ ;  $p = .0003$ ), Word-Meanings-in-Context ( $d = 0.20$ ;  $p = .0227$ ), and Target Word

Association ( $d = 0.15$ ;  $p = 0.0830$ ). Regarding the standardized measures, the effect of treatment on Gates–MacGinitie Reading Comprehension was relatively small ( $d = 0.15$ ;  $p = .0568$ ), whereas the effect of treatment on SAT-10 Reading Vocabulary was very small and nonsignificant ( $d = 0.005$ ;  $p = .5045$ ). It is worth noting that the three-level model for Target Word Association encountered convergence problems, so two-level models accounting for the nesting of students within classes were fitted, and bootstrapped standard errors were estimated and adjusted to account for the additional clustering within schools.<sup>6</sup>

To investigate whether any of the effects of treatment were greater for LM learners or NE speakers, a statistical interaction between LM learner status and treatment was included in models for each outcome. Results indicated that the effect of treatment did not differ significantly by LM learner status for any of the six outcomes. To provide further context for understanding this result, Table 6 presents the sample pretest means and the adjusted posttest means for each condition, separately for LM learners and NE speakers. The LM learners in each condition performed lower than their NE speaker classmates at both pretest and posttest on all measures; as shown in the far right column in Table 6, these differences were roughly one half of one pooled sample standard deviation at posttest ( $d$  ranged from 0.39 to 0.61). Similarly, the effects of treatment were not found to differ substantially as a function of students' pretest achievement. For each posttest measure,

**Table 5. Results of Multilevel Models Evaluating the Effects of Treatment on each Student Outcome (N = 476)**

		Target Word Mastery	Morphological Decomposition	Target Word Association	Word-meanings-in-Context Comprehension	Gates Reading Comprehension	SAT-10 Reading Vocabulary
Fixed effects	Intercept	21.930***	13.062***	36.618***	12.999***	495.782***	644.165***
	Treatment (in Cohen's $d$ )	0.391***	0.217***	0.154	0.200*	0.146	0.005
Variance components	School (Intercept)	0.617***	0.061***	<sup>a</sup>	0.321***	10.764***	12.150***
	Class (Intercept)	0.095**	0.005	0.417	0.265***	7.580***	8.075***
	Child (Residual)	16.694***	4.563***	3.248	10.100***	433.415***	508.511***
Goodness of fit	-2Log-Likelihood	2701.4	2078.3		2467.3	4253.4	4328.9
	AIC	2733.4	2110.3		2499.3	4285.4	4360.9
	BIC	2800.1	2177.0		2565.9	4352.0	4427.5
Likelihood ratio test	$H_0$ : Treatment = 0	27.840***	13.130***	2.870	5.190*	3.630	0.0010

Note: All models include pretest scores, ethnicity, and language minority learner status as covariates.

<sup>a</sup>The three-level model for this outcome encountered convergence problems. Instead, a two-level model with random effects for classrooms was fitted and the standard errors were adjusted for the clustering within schools using bootstrapping methods; although this approach does not allow for the direct estimation of a variance component for school, it nonetheless accounts for the nesting of students within class within school in conducting the test for the statistical significance of the treatment effect.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

**Table 6. Results on Student Outcome Measures for Language Minority (LM) Learners and Native English (NE) Speakers, by Treatment and Control, expressed as Sample Means at Pretest and as Fitted Means at Posttest, Adjusted for Pretest Achievement and Demographic Covariates based on Multilevel Models ( $N = 476$ )**

Outcome measure	Pretest				Posttest				
	Control sample mean		Treatment sample mean		Control adjusted mean		Treatment adjusted mean		LM-NE difference (in Cohen's $d$ )
	LM	NE	LM	NE	LM	NE	LM	NE	
Target Word Mastery (raw score)	19.78	23.83	17.93	20.78	20.78	23.85	23.18	26.26	0.51
Morphological Decomposition (raw score)	11.11	13.09	10.83	13.09	12.48	13.97	13.28	14.78	0.41
Target Word Association (raw score)	34.12 <sup>a</sup>	35.98 <sup>a</sup>	32.11 <sup>a</sup>	34.63 <sup>a</sup>	35.14	37.85	35.85	38.57	0.60
Word-Meanings-in-Context (raw score)					12.42	14.33	13.36	15.27	0.39
Gates-MacGinitie Reading Comprehension (extended scale score)	498.72	520.61	491.66	505.76	491.91	505.32	496.74	510.15	0.42
SAT-10 Reading Vocabulary (developmental scale score)	633.19	651.48	622.39	649.92	638.06	659.14	638.25	659.33	0.62

<sup>a</sup>The pretest for the Target Word Association was conducted one month prior to the other pretests with a slightly different sub-sample of 405 participants, and was therefore not included as a covariate in estimating treatment effects. As a result, these estimates should be interpreted with caution.

a term for the interaction between treatment and the given measure at pretest was tested (in the case of the two measures not given at pretest, interactions with the standardized pretest measures were used instead). Each of these interactions were found to be nonsignificant, with the exception of a small, but significant interaction for the Word Mastery outcome ( $d = -0.04$  for each additional point on the pretest;  $p = .0002$ ), which suggested that students with lower initial knowledge of the taught words benefited slightly more (e.g., a student who knew 5 fewer words than an average student learned approximately 1 additional word from the intervention).

Given the quasi-experimental nature of the study, one threat to the validity of causal inferences about these effects of treatment is that the treatment teachers may have differed from the control teachers in the quality of their instruction in ways unrelated to the treatment, and that these differences may have an impact on results for student achievement that would bias the effect of the treatment effect. Comparison of the instructional quality ratings of teachers in the treatment and control classrooms indicated that the instruction provided by the two groups differed only in ways that were related to the ALIAS instructional approach. As described previously, observers used a three-point scale

(0 = low, 1 = medium, 2 = high) to rate teachers on nine quality indicators that were specific to the instructional approach in the intervention but observable in any class (e.g., Teacher facilitates student talk) and six indicators of instructional quality that were not specific to the intervention (e.g., Teacher responds effectively to misbehavior); averaging these two groups of indicators separately yielded two composite values for each teacher. Table 7 provides the mean ratings for these two composites as well as the individual indicators that comprise them, separately for treatment and control teachers. Treatment and control teachers did not differ substantially on the non-ALIAS instructional quality composite ( $d = -0.08$ ;  $p = .8566$ ), however the treatment teachers were more than a standard deviation higher than the control teachers on the ALIAS-specific instructional quality composite ( $d = 1.12$ ;  $p = .0383$ ). These findings provide evidence that the treatment teachers did not differ from the control teachers in the quality of their instruction in ways unrelated to the treatment and thus support the assumption that the differences in instruction provided by the treatment and control teachers was largely the result of the ALIAS program rather than preexisting differences between the two groups of teachers in their instructional approaches.

**Table 7. Comparison of Treatment and Control Teachers on Composites and Individual Items from Instructional Quality Indicators during Intervention Period on a Three-Point Scale (0 = Low, 1 = Medium, 2 = High)**

Instructional quality indicator		Treatment teachers (n = 12)	Control teachers (n = 6)	Mean difference (in Cohen's <i>d</i> )
Non-ALIAS-specific elements	Non-ALIAS-specific Composite (alpha = .85)	1.40	1.43	-0.08
	H1: Teacher is prepared for class	1.59	1.79	
	H3: Teacher manages students to maximize time on task	1.25	1.32	
	H4: Teacher responds effectively to mis-behavior	1.41	1.31	
	H5: Teacher demonstrates commitment to and belief in their lesson	1.44	1.43	
	H7: Teacher keeps instructional materials organized	1.61	1.63	
	H13: Students demonstrate high levels of engagement	1.10	1.10	
ALIAS-specific elements	ALIAS-specific Composite (alpha = .77)	0.89	0.65	1.12*
	H2: Teacher effectively leads classroom discussion	1.25	1.32	
	H6: Teacher affirms correct word definitions and usages	1.22	0.45	
	H8: Classroom includes visible vocabulary-related information	1.30	0.88	
	H9: Teacher regularly repeats words and word meanings in different contexts	0.83	0.39	
	H10: Teacher uses personal anecdotes and examples to teach words	0.52	0.31	
	H11: Teacher supports students' writing	0.58	0.47	
	H12: Teacher reviews previously taught words	0.15	0.00	
H14: Teacher facilitates student talk	0.89	0.81		
H15: Teacher responds to students' needs	1.27	1.25		

\* $p < .05$ .

### ***Ease of Implementation***

Our second research question addressed the implementation of the academic vocabulary program. Our conception of implementation was twofold: the actual uptake of the program in the classrooms studied, as measured by fidelity of implementation, as well as teachers' insights into the ease of implementation of the program. In what follows, we first describe findings on the reported and observed measures of fidelity of implementation and then report on three themes that emerged from the interview and survey data concerning teachers' perceptions of program implementation.

### **Fidelity of Implementation**

Overall, descriptive analyses of data from classroom observations and weekly log data on lessons completed

and minutes of instructional time for the treatment teachers indicated that the treatment curriculum was implemented with good fidelity.

### ***Reported Fidelity***

Reported fidelity of implementation of the treatment was high, as indicated by the self-report data from the treatment teachers' weekly logs. The top two rows of Table 8 provide summary statistics for the percentage of daily lesson components reported as implemented and the reported instructional time spent on each lesson. As shown, teachers reported implementing an average of 86.9% of the essential lesson components for the 72 daily lessons and reported spending an average of 52.5 minutes per day on ALIAS lessons. Although there was some between-teacher variation in the proportion of components implemented and minutes spent on ALIAS,

**Table 8. ALIAS Implementation Data from Weekly Logs and Observations (Teachers  $N = 12$ ; Lessons Reported on Per Teacher  $N = 72$ ; Total Lessons Observed,  $N = 75$ )**

		Mean	SD	Min	Max
Self-reported	Lesson components implemented	86.9%	6.6%	76.7%	95.4%
	Instructional minutes per day	52.5	9.5	43.5	76.8
Observed	Lesson components implemented with fidelity	73.6%	14.6%	50.0%	100.0%

even the teacher with the lowest fidelity nonetheless reported implementing a high proportion (over 75%) of the lesson components and similarly, no teacher reported substantially less time spent than the allocated 45 minutes per lesson.

### **Observed Fidelity**

Data from observations in the treatment classrooms largely confirmed that average fidelity was high across all lessons. The bottom row of Table 8 displays the descriptive statistics for the percentage of lesson components rated as implemented with good or excellent fidelity. As shown, the average teacher implemented 73.6% of the components across the seven lessons observed with good or excellent fidelity. As might be expected, variation in observed fidelity between teachers was greater (ranging from 50% to 100%) than that for reported fidelity.

### **Teachers' Perceptions of Program Implementation**

Three overall themes emerged from an analysis of the interview and survey data focused on gaining insight into ease of implementation: 1) The challenge of meeting students' instructional needs, 2) the central role of appropriate and appealing text in promoting classroom talk and vocabulary learning, and 3) the role of high-quality curricula and professional networks in building teachers' instructional capacity (i.e., their ability to implement the program effectively as well as draw lessons for their practice beyond the program).

#### ***The Challenge of Meeting Students' Instructional Needs***

In weekly log entries and in the interviews, teachers commented on how ALIAS addressed their students' instructional needs in spite of what appeared to be substantial challenges at the outset. Initially, teachers perceived that many ALIAS instructional activities were too difficult for their students. In log entries they described

students as having had minimal prior practice understanding and using academic vocabulary ("the students are having a really hard time with the word *welfare*," "they struggled with the word *classify*, having no clue what it meant"). However, in the end-of-program interviews, the majority of teachers cited the benefits of the program and setting high expectations. As one teacher explained in a typical response, "It was very rigorous for my kids...[S]tarting real high—the expectations are really high of these kids—that was very helpful, because by the end of Unit 8, they met those expectations."

In particular, although writing activities took a long time to complete, log entries suggested that teachers believed that the ALIAS writing instruction resulted in increased writing competence ("I think this part of the program has been extremely helpful for their writing," "This writing activity always takes a long time, however, we are building some very good writers," "I have seen tremendous growth this year with my students' writing. They are so independent. I am looking at their work from the beginning till now and I am impressed!"). This theme was less prevalent in the interviews, which focused more on the macro-level process of implementing ALIAS as opposed to specifics about the program; however, unprompted, three teachers singled out improvement in their students' writing, while a fourth teacher suggested the value of having more extensive writing (e.g., essays instead of paragraphs). Although the response to the structured writing instruction was positive, the majority of the teachers felt that the peer revision task was not a good use of instructional time; they believed their students did not have the skills necessary to support one another's revision efforts.

#### ***The Central Role of Appropriate and Appealing Text in Promoting Classroom Talk and Vocabulary Learning***

In weekly logs and in interviews, teachers also highlighted the role of the texts in promoting valuable discussion. Each ALIAS unit was designed to begin with a

short text chosen to contain big ideas worthy of discussion, in the interest of promoting rich classroom talk. Beginning with discussions before and after reading the article on each Day 1, the classroom conversations were meant to clarify key concepts in the text and tasks and elicit personal responses from students, while allowing for additional practice hearing and using academic language. Weekly logs reported that the articles engaged students (“this is a high-interest article,” “my students were fascinated by the idea of gender schools”), and were the basis for lively discussions (“they really like to debate this issue,” “very good day—excellent discussion about discrimination and how students are affected both now and in the past,” “Students really enjoyed this article. Pair work to generate solutions to cyber-bullying was fun for them, and they liked giving answers that later went on the chart”). These comments suggested that the teachers were successful in using the provided texts as platforms for improved student discussions.

This idea of starting with engaging text to design instruction that encourages critical thinking, lively discussions, and writing practice was one of the key understandings for many of the participating teachers. When asked what they would take away from the experience, almost half of the interviewed teachers cited the idea of beginning with high-quality text to design their vocabulary instruction. As one teacher explained, “There’s definitely things I’m gonna take from this—spending time with a certain group of words, pulling them out of the articles...definitely the basic concept of giving the context, but teaching explicitly to those words.” During the interviews, two teachers also indicated that ALIAS, with its emphasis on discussion and writing, encouraged broader literacy practices than those traditionally associated with vocabulary instruction. As one teacher put it, “[T]he vocabulary instruction—it’s not just the vocabulary. And I’ve tried to explain this to several other people—the vocabulary is one element of it.”

Although teachers initially expressed concerns about the time required by the program in the early logs, these declined over time and only one teacher expressly raised this issue in the exit interviews. Four teachers did express some concern about omitting standards-based instruction that students would be expected to have had for the state test, while one teacher also regretted having to cut time spent on sustained silent reading. Yet even among this group of teachers, there was some tension between the value teachers saw in the ALIAS program and the pressure they felt to cover other material. As one teacher explained,

It took more time than I thought it would. And maybe not so much—maybe I did think it was gonna take more time—but more time than I would have liked. And I don’t even know about that—maybe I thought the time was well spent. But I guess I should say it’s more time than I have.

In the teacher logs and interviews, some teachers suggested better integrating the ALIAS curriculum and the requirements of the district standards as a solution to this problem (“I’m having difficulty finding time to teach my students the academic words for the STAR Test. In the future the program may want to include these words so two functions could be fulfilled in the same lessons,” “If ALIAS was gonna go again, which I’m sure it will, it needs to really be looked at and integrated with standards”).

### ***The Role of High-Quality Curriculum and Professional Networks in Building Teachers’ Instructional Capacity***

The logs and interviews revealed three types of support important to teachers’ use of the program and their reported improvement in teaching vocabulary: the program curriculum itself, other teachers in their school using the program, and the monthly meeting with the Program Specialist.

First and foremost, we received a strong message that program materials matter. These materials were discussed repeatedly in the teacher logs and the interviews as the most supportive element to doing this work. During the end-of-year interview, when asked to choose the essential component from a list of all forms of ALIAS support, one teacher answered, “the curriculum. Because when it comes down to it, it’s you, the classroom, and the curriculum. And so that’s the biggest piece of support.” The majority of teachers interviewed described the ALIAS materials as highly supportive and self-explanatory. As one teacher stated during the interview, “Honestly, the fact that I had the binder with the scripted lesson and the outline—that was all the support I needed. I didn’t need anything else.” Moreover, some teachers noted in their logs that the information increased their own learning. One teacher, for example, explained how the ALIAS curriculum boosted her knowledge base: “To be honest, I always get affect and effect mixed up. This lesson actually helped me with it.”

Related to the program materials, during the interviews, teachers singled out the curriculum’s optional sample instructional models (which provided scripted language for instruction) as a positive feature; 8 out of the 11 teachers indicated that having the optional script was beneficial, although they did not always use it. Several teachers indicated that they thought this type of instructional model was most beneficial to new teachers; teachers who reported that they relied heavily on the models during the early stages of implementation and later used them as a reference, picking and choosing ideas and examples were, indeed, relatively new (less than four years experience). As one teacher reported,

Quite frankly, I started out using all the scripted lessons. But as I got more comfortable, then I would do my own or change a little bit here and there, you know, change it around. Because you gotta have that teacher comfortableness with it.

Other adults were also reported to play an important role in teachers' use of the program. Although formal collaboration was not required, most teachers had access to one or two other teachers in the school also using the program, and six of the teachers indicated that they took advantage of this opportunity. All of these teachers spoke favorably about their collaborations, which largely centered on answering the day-to-day question of, "How do I do this?" or, as one teacher put it, "debugging" the program. Discussions included sharing ideas for classroom arrangements and supplemental materials.

Strikingly, teachers' reactions to the local Program Specialist were more varied than their responses to the other supports. While three teachers appreciated the Program Specialist's support, describing her role as helpful for clarifying things promptly ("great support and quick answers to ALIAS questions from [the Program Specialist]"), three others indicated neutrality ("[The Program Specialist] had lots of constructive things to say and stuff like that, but it wasn't like I felt like I needed that as a teacher"), and four expressed a lack of satisfaction ("I don't think she knew the daily program. I think she came to watch management...But it didn't really help me teach the program."). In part, these reactions were associated with different understandings of the Program Specialist's role. The teachers who appreciated the support of the Program Specialist believed that her role was to help teachers. In contrast, those who were less satisfied perceived the Program Specialist to be playing an evaluative, rather than supportive role. The level of satisfaction with the Program Specialist also varied systematically by teachers' years of experience; in the extremes, the three teachers who were highly satisfied with the Program Specialist had all taught for five or fewer years, while the four teachers who were dissatisfied had all taught for six or more years. The existence of other professional networks may have also played a role in teachers' evaluation of the Program Specialist; the teachers most enthusiastic about this support reported little opportunity to collaborate with colleagues, while some of those most opposed were already working with other teachers. Overall, these varied reactions to the Program Specialist highlight the importance of considering contextual factors and building in flexibility when designing program supports.

## ***ALIAS Instruction in Contrast With Standard Practice***

To further shed light on the middle school English language arts (ELA) context, our final research question focused on the degree to which the implemented instruction contrasted with standard instructional practice. As previously noted, data from observations in the treatment and control classrooms demonstrated that the instruction that students received in the two conditions differed qualitatively in the prevalence and nature of vocabulary instruction. In this section, we report results from quantitative analyses on the differences between the curricular content observed in the treatment and control classrooms, as well as results from the teacher survey, which provide some insight into what the ALIAS instruction replaced in their classrooms.

### **Curricular Content and Instruction in Control Classrooms**

Observations in the control classrooms indicated that the curriculum content taught in these classrooms differed substantially from that observed in the treatment classrooms, with very little instruction in the control classrooms focused on vocabulary. Table 9 provides descriptive data on the curricular focus observed in the control classrooms, as a proportion of the 15-minute intervals observed across the seven control teachers. As shown, vocabulary was the primary focus on instruction in only 10.3% of the 87 intervals observed. The most common curricular focus in the control classrooms was literary analysis and related comprehension activities with literary texts; this was the focus for approximately 40% of the instructional time observed. Reading comprehension with informational text and writing were also more common curricular foci than vocabulary. In contrast, every lesson observed in the treatment classrooms during the intervention period had vocabulary as its curricular focus; even in lessons demonstrating low fidelity to the curriculum, vocabulary and related skills (e.g., morphological awareness, writing with target vocabulary) were the focus of instruction.

### **Survey Data About Standard Practice**

Data from the teacher survey generally supported the claim that their instruction changed in ways consistent with the ALIAS approach; for instance, a majority of treatment teachers reported that the treatment period included more time for instruction in various aspects of vocabulary, for collaborative learning in pairs, for students to explain and reflect on their thinking, for students to focus on word choice in writing, for guided reading, and for use of graphic organizers.

In addition, the teacher survey data shed light on the curricular content and instruction that ALIAS replaced.

**Table 9. Curricular Content Focus in Control Classrooms (*n* Teachers = 7; *n* Observations = 29; *n* Intervals = 87)**

Curricular content	Percent of 15-minute intervals observed ( <i>n</i> of intervals)
Literary response; analysis and comprehension of literary texts	40.2% (35)
Reading comprehension of informational texts	12.6% (11)
Writing (composition)	11.5% (10)
Writing (conventions and grammar)	10.3% (9)
Vocabulary	10.3% (9)
Test-taking or test-taking strategies	8.1% (7)
Listening and speaking	3.5% (3)
Independent reading	2.3% (2)
Classroom management, discipline, or other (non-instructional)	1.2% (1)
Decoding and fluency	0% (0)

A majority of treatment teachers reported spending less time during the treatment period on four elements that they had previously taught frequently (i.e., 3–4 times a week): narrative elements (e.g., characters, setting, plot); explicit reading comprehension strategies and skills (e.g., predicting, summarizing, identifying main ideas); independent reading; and reading aloud to students. A majority of the control teachers also reported teaching each of these elements with high frequency (3–4 times a week) or moderate frequency (2–3 times a week), supporting the claim that these were the most common elements replaced by ALIAS.

A majority of treatment teachers also reported spending less time on several elements that they had previously taught with moderate frequency: identifying text features (e.g., index, headings); analyzing graphics; oral reading speed and accuracy; several aspects of writing, including writing narratives, organizing writing, and setting a purpose for writing; and independent writing of reports or expressive writing (e.g., journals).

## Discussion

The primary goal of the present study was to generate new insights about vocabulary instruction by evaluating the effects of a vocabulary program designed for use in urban middle schools with high numbers of language minority learners, in an effort to bolster students' vocabulary and reading comprehension skills. A secondary goal was to investigate its ease of implementation—defined by its use in these mainstream classrooms and teachers' perceptions of the program—given that the ultimate success of any program depends upon its utility in engaging teachers and students from a wide variety of backgrounds (Donovan et al., 2003; Schneider, 2004).

The present study makes a substantive contribution to research and theory in several ways.

If literacy research is going to inform instructional improvement at scale, it must be conducted in ecologically valid ways. Specifically, there is a need for studies that test classroom-based interventions that feature a comprehensive, multifaceted approach to vocabulary instruction, in contrast to focusing on only one aspect of vocabulary knowledge. This approach must be one that can be readily adopted by teachers concerned with the academic needs and engagement of their struggling students. In addition, given the linguistic diversity inherent in today's school-aged population, interventions should be designed with language minority learners in mind, but should also be appropriate for their monolingual classmates, not carrying any hidden costs for this group of students. Overall, the findings show promise in developing effective multifaceted vocabulary instruction for successful implementation by ELA teachers in middle school classrooms with high numbers of language minority learners. At the same time, and consistent with previous research, the findings serve to highlight and reinforce the challenge of bolstering vocabulary and reading comprehension skills for students who are not on grade level. What follows is a discussion of the findings and their corresponding implications for research and practice.

Specifically, significant program effects were evident on researcher-developed vocabulary measures targeting knowledge of the words taught, knowledge of word meanings in context, and morphological skills, with effect sizes consistent with previous research on teacher-implemented classroom-based instruction. For instance, the effect on knowledge of words taught ( $d = 0.39$ ) was comparable to the effect size ( $d = 0.34$ ) found by Carlo et al. (2004) and the effect sizes across the measures were consistent with those found for

large-group educational interventions (see Bloom, Hill, Black, & Lipsey, 2008). While many studies do not include standardized measures of vocabulary or comprehension, we so included these in an attempt to detect transfer effects on measures outside of those based on the curriculum (Pearson et al., 2007). The effect of treatment on a standardized measure of reading comprehension was marginally significant and small in magnitude, as discussed below. However, the program did not show significant or practically meaningful effects on a standardized measure of reading vocabulary. Although more research is needed to address this latter finding, it may be, in part, explained by the fact that the words used in this measure are sampled from a different domain as compared with that of the academic vocabulary words that were the focus of the teaching in this study. Finally, although implemented in urban classrooms with high numbers of language minority learners and designed with this population in mind, we found that the curriculum was equally beneficial for their monolingual English-speaking classmates.

There are several considerations for researchers and practitioners that emerge from this instructional study. First, the findings suggest that text-based academic vocabulary teaching is a promising approach for improving early adolescents' vocabulary and comprehension. That is, starting with a short piece of accessible grade-level text and identifying key academic words—words that appear in textbooks across the disciplines—to teach deep word knowledge, is an approach that has not been sufficiently tested in research with older students from diverse backgrounds but that proved a step in the right direction for a large number of struggling readers. Without careful selection of text, these students would likely get bogged down with other linguistic features of typical textbook passages (e.g., syntax or voice; for a discussion see Schleppegrell, 2004) or the idiosyncratic demands in background knowledge required by content area texts (Fang & Schleppegrell, 2008; Schleppegrell, 2004). Thus, in designing this program, we identified expository texts with accessible content relevant to youth and then identified the corresponding, high-utility academic words (most often abstract concepts) that appeared in the articles, which became a platform for vocabulary instruction.

Moreover, the study and its findings extend previous research on vocabulary instruction into the middle school context. Specifically, it suggests that vocabulary at the middle school is greater than the sum of the parts; that is, a number of lessons using different modalities to learn words deeply—lessons that move beyond simple definitions to focus on building depth of word knowledge (multiple meanings, morphological analysis) over time show promise in bolstering vocabulary and comprehension skills of the middle schooler.

Given that word knowledge is thought to develop incrementally with each meaningful, contextualized encounter with the word (Beck et al., 2002; Stahl & Nagy, 2006), the lesson cycle within each unit provided multiple, planned exposures to the words in authentic and natural ways. Consistent with current research on the multidimensional nature of word knowledge, some lessons included a focus on definitions and semantically related words, others focused on the polysemous nature of many of the target words, while still other lessons focused on their transformation into other morphological forms (Nagy & Scott, 2000; Stahl, 1999).

The design of the multifaceted curriculum presents a departure from standard practice in the participating classrooms; the approach is in sharp contrast to the common practice of starting with a list of words, memorizing definitions, and completing basic activities (e.g., using the words in disconnected sentences) or using words that publishers provide during textbook work, which are not always high-impact academic words that are required for comprehending a range of texts. The texts we selected were replete with words that have been found to also appear frequently in the textbooks (Coxhead, 2000; Nair, 2007). In addition to reading and referring to the text, students were engaged in significant amounts of structured, academic talk and writing, in notably greater proportion than in the traditional instruction observed in the control classrooms.

Finally, we confirmed that vocabulary instructional approaches and activities deemed effective in research with monolingual English speakers (e.g., Beck et al., 2002; Graves, 2000, 2006; Stahl & Nagy, 2006) could be effective with learners from diverse linguistic backgrounds. To design effective instructional environments for *all* learners, particularly given the increasing diversity of the school population, we must determine which strategies are effective at the classroom level. Consistent with the findings of Carlo et al. (2004), who conducted their study in fifth-grade classrooms with learners from diverse linguistic backgrounds, we found no differences in treatment effects for language minority learners and their monolingual English-speaking peers. Given the mean achievement level at pretest in the participating classrooms—the 35th percentile in reading comprehension—rigorous instruction to meet the needs of both groups is important and pressing.

However, the equal treatment effects for the two groups raises questions about how to narrow the achievement gap. Consistent with other studies that have investigated treatment by language background interactions (e.g., Carlo et al., 2004; Gunn, Biglan, Smolkowski, & Ary, 2000) as well as recent trends on large-scale assessment (e.g., NAEP long-term trend study; NCES, 2008), the LM learners in the current study demonstrated as much, but not more, improvement than their NE

speaker classmates. As a result, they remained approximately half a standard deviation below their peers in vocabulary and reading comprehension at posttest. To address the concerns of educators who seek to reduce achievement differences, further research is needed to determine the conditions under which teachers can accelerate LM learners' growth such that they can catch up with their monolingual peers. It may be that instructional techniques different than those that have been investigated to date are necessary or that the instructional focus (e.g., the words targeted for instruction) needs to be different for LM learners. On the other hand, it may be that the same instructional techniques and foci should be provided to LM learners, but in greater dosage (i.e., more instructional time) or greater intensity (e.g., in smaller or more individualized settings).

The small effect on students' performance on a standardized measure of reading comprehension raises several important issues. Although marginally significant, the magnitude of this effect size ( $d = 0.15$ ) was comparable or slightly larger than that found by Carlo and colleagues (2004) on a researcher-developed measure ( $d = 0.08$ ). In more practical terms, this effect translated into an *additional* 8 to 9 months of growth, compared with normative data on growth in the sixth-grade academic year (Bloom et al., 2008). Although improving reading comprehension performance is indeed the ultimate goal of many vocabulary interventions, this relationship has not been demonstrated prominently in the existing research.

An underlying theory of much vocabulary instruction is that increasing word knowledge and word-learning strategies will, over time, benefit the reader by increasing the ability to make meaning from text. For example, this may occur by decreasing the ratio of unknown words encountered in text or by increasing students' metacognitive ability to derive the meaning of unfamiliar words. However, those vocabulary studies that have shown effects on reading comprehension measures have typically done so using a researcher-developed measure derived from the very same texts with which the words were taught (NICHD, 2000)—a local measure of near transfer (Pearson et al., 2007). In contrast, it has been much more difficult to show effects on what researchers call far transfer or global measures, such as the standardized, norm-referenced measure used in the present study. For example, the National Reading Panel Report (NICHD, 2000) found only two studies that showed that effects from vocabulary instruction transferred to a norm-referenced comprehension measure.

There are several possibilities for explaining this promising finding. First, it is possible that this effect was due to the work with text in the program. As noted, short pieces of expository text were the basis of the

program; to teach the target academic words we begin with the content of the text. Four days of the eight-day cycle involve students working with the article—revisiting it, reading for meaning, and discussing its content to build up word knowledge, and perhaps the in-depth focus on a single text improved overall reading comprehension. Second, in considering the program effects on reading comprehension, we also highlight that the program had significant effects on morphological awareness skills—our largest effect size after that associated with words taught. Morphological awareness, as a facet of metalinguistic awareness, has been identified as an area of weakness for language minority learners and for poor readers, yet has a strong, predictive relationship with reading comprehension performance for these populations (Carlisle, 2000; Kieffer & Lesaux, 2008; Nagy, Berninger, & Abbott, 2006). Finally, the results may be further understood by examining typical ELA instruction in the control classrooms; on the standardized comprehension measure, students in the control classrooms showed no growth in absolute terms and lost ground against national norms over the course of the academic year. Although control teachers reported vocabulary as their students' primary weakness and instructional need, our observations in these classrooms revealed only approximately 10% of instructional time could be classified as vocabulary instruction. What instruction there was tended to be incidental, i.e., focused on the rare, unfamiliar words (e.g., *cannibal*, *azure*, *slurp*) in a given passage, and superficial, i.e., characterized by providing a single definition or example for a given word without opportunities for processing or practice with the word meaning. Moreover, in these classrooms, the majority of instructional time focused on literary analysis, and only approximately 10% of time was focused on instruction in reading comprehension skills. Although not directly addressed in this study, the focus of standard ELA instruction in these control classrooms is likely a function of classroom materials—novels and literature anthologies—coupled with the background in literature and drama that many of these teachers report.

Following on the findings that the program proved promising as compared with standard practice in bolstering vocabulary and comprehension skills, we investigated teachers' perceptions of the program. This line of investigation was based upon the premise that establishing a program's effectiveness via empirical data showing its positive effects on student achievement is a necessary but not sufficient endeavor, if research is to influence instructional improvement. This is particularly important in the context of the middle school where reading instruction plays only a minor role in the overall curriculum. Three particular themes emerged from an analysis of relevant data, reinforcing and shedding light on several findings previously discussed: the

challenge that teachers experience in meeting students' instructional needs, the role of text in promoting vocabulary and classroom talk, and the role of curricula and implementation support in building teachers' capacity for effective literacy instruction.

Initially, teachers perceived that many ALIAS instructional activities—especially the writing activities—were too difficult for their students. Many described students as having had minimal prior practice understanding and using academic vocabulary and found it challenging to implement the program at the outset. However, at the end of the program, the majority of teachers cited the benefits of the program and its role in setting high expectations for student language learning. In addition, teachers and students alike reported the engaging nature of the articles, with many teachers commenting specifically on the role that the text played in this vocabulary instruction; when asked what they would take away from the experience, almost half of the interviewed teachers cited the idea of beginning with high-quality text to design their vocabulary instruction. Also related to ease of implementation, there was an unequivocal message from teachers that materials matter when instituting instructional change. In this case, the materials were seen as the primary strength of the program, considered by participating teachers to be straightforward yet detailed. Discussions of these effective program materials simultaneously reinforced the lack of a one-size-fits-all approach to supporting teachers; the materials provided teachers with the option of differing levels of support (i.e., using a one-page lesson outline or a longer sample script), which turned out to be important according to many teachers.

Strikingly, although those teachers with colleagues similarly using the program found the informal collaborative opportunities helpful, teachers' impression of the support provided by the Program Specialist—a teacher with considerable experience in the district who knew the student population and the program well—were varied. For some teachers, particularly those with fewer than 5 years of teaching experience and those without colleagues in their school using the program, meetings with the Program Specialist were perceived as an excellent source of support and an important outlet for discussing their instruction. For others, this was not the case. Although the challenges and complexities of effectively supporting teachers to improve practice have been well-documented (Bryk & Schneider, 2002; Elmore, 2004; Guskey, 1989, 2002; Marsh et al., 2008), that the same adult can be perceived as helpful and essential by one teacher and as evaluative or nonessential by another is a cautionary tale for middle school literacy reform efforts, especially as the literacy coaching model gains momentum across the nation (Hall, 2004; Marsh et al., 2008).

---

## Implications and Directions for Future Research

As is most often the case, the findings from this study raise further questions that warrant investigation, to continue to advance knowledge in this underdeveloped area of research. First, the study was designed as a sound quasi-experimental evaluation (i.e., meeting the standards of best evidence articulated by Slavin et al., 2008) to determine the effects of the program across a heterogeneous sample of teachers and students in urban middle schools. The lack of differences between teachers in the control and treatment classrooms on key observed variables, coupled with the positive program effects, led us to conclude this multicomponent curriculum is a promising approach for the mainstream, urban classroom with high numbers of language minority learners. Nevertheless, in order to confirm causal inferences about the program's effects on students' vocabulary and reading comprehension skills, there is a need for an evaluation using random assignment. Given evidence that some effects are small and that implementation may vary by context, such research needs to be at a sufficiently large scale to observe these effects in typical conditions. Simultaneously, this research must continue to advance a rich and contextualized understanding of literacy reform beyond the development of effective, empirically-based strategies and instructional approaches, to focus on ease of implementation; factors that influence the middle school teacher's buy-in, uptake, and sustained use of any literacy program or approach; and effective professional development to increase teachers' capacity to meet the needs of their struggling readers.

Second, in conducting the study, we were guided by the goal of designing effective and engaging classroom-based vocabulary instruction with high ecological validity, particularly given the context of the urban middle school. The trade-off for doing so, from a research perspective, is that the academic vocabulary program studied is multifaceted in nature—using multiple techniques and instructional components to build up word knowledge and word-learning strategies—and we are not able to isolate which strategies are most effective in advancing students' knowledge. Further research should use multiple treatment conditions to identify the active ingredients in the approach.

Third, although vocabulary instruction research has not typically assessed students' writing, in a subsequent study we would include a measure of students' writing abilities to capture potential program effects in this domain; in the end-of-program interviews, all teachers reported effects on their students' writing skills. While we included writing activities in each unit, we anticipated this would complement and

supplement regular ELA instruction. However, our observations of standard practice revealed that only approximately 10% of instructional time was spent on extended writing, suggesting that the quantity and quality of writing activities may be involved in the effectiveness and perceived usefulness of vocabulary instruction. Although we did demonstrate short-term program effects on specific measures of vocabulary and reading comprehension administered immediately after the 18-week intervention, further research is needed to determine whether there are long-term effects of the program. In addition, findings do raise questions about dosage and duration—it may well be that a program such as this provided for the entire academic year would show stronger effects.

Finally, the design of this program, coupled with the push toward more vocabulary instruction in classrooms, raises questions and speculations about the design of this instruction across grade levels. With students' academic achievement in mind, it would likely be better if the text-based instruction in academic vocabulary words such as *evidence* and *method* was commonplace in the upper elementary years, to predispose students to more comprehension with the linguistically demanding textbooks they will encounter in middle school and in turn, to lay a foundation for increasingly discipline-specific vocabulary learning. These high-utility academic words might be conceptualized as to transition to the content words that correspond with the increasingly specialized and sophisticated conceptual knowledge that students must master in the middle and high school years. Under these circumstances, a program such as this one might function as an intervention for a smaller number of students in the middle school rather than as a part of the regular ELA curriculum for classrooms with high numbers of strugglers. Alternatively, this may be the right strategy for middle school ELA, if complemented by discipline-specific vocabulary instruction in the content area classrooms. Future studies should seek to identify the appropriate context and mix of language instruction that will yield the greatest outcomes for students at different developmental levels.

Although future research in this area is clearly needed, these findings highlight the promise of improving academic vocabulary instruction as a key ingredient in increasing opportunities to learn for students in urban middle schools. As educators and policymakers seek to improve the literacy outcomes of learners from linguistically diverse backgrounds, they must attend not only to students' specific sources of difficulties presented, but also to the needs of teachers in urban contexts for curricula and professional networks that support their improving practice.

## Notes

<sup>1</sup> In one school, all participating teachers delivered the intervention to guard against contamination effects; these three teachers had extensive common planning time and a long history of collaborating together to plan instruction.

<sup>2</sup> As noted above, this measure was not included in the pretest battery. Rather, it was piloted with a subsample of 17 of the 21 participating classrooms one month prior to the pretest. This reliability estimate comes from this pilot. Unlike pretest scores for the other measures, students' scores on this pilot were not used as covariates in the multilevel models because this would have necessitated eliminating four classrooms from the analyses; however, the results reported were largely the same for a model fitted to this subsample that did include this pilot score as a covariate.

<sup>3</sup> Due to technical malfunction, the interview for the 12th participating treatment teacher was not taped with clarity and thus not used in the analyses.

<sup>4</sup> We recognize that this portion of our study is based on a small, purposive sample; we use the findings strictly to begin to explore the ways in which teachers' perceptions and reported experiences might mediate program implementation.

<sup>5</sup> Attempts to investigate the variation in treatment effects at the teacher and school level using random effects led to convergence problems, and alternate models (e.g., with treatment by school interactions specified as fixed effects) indicated that there was limited variation in treatment effects across this relatively small number of teachers and schools.

<sup>6</sup> Although this alternative model has the disadvantage of not explicitly modeling variation at the school level as well as at the class level, it nonetheless accounts for the nesting of students within classes within schools and thus provides a conservative and appropriate test of the null hypothesis that the effect of treatment is zero in the population.

This research was supported by grants from the William T. Grant Foundation and the William and Flora Hewlett Foundation awarded to Nonie K. Lesaux. The authors would like to thank Carol Barry, Teresa Walter, Jennifer Cheatham, Amy Griffiths, Phoebe Sloane, Cara Wells, Tisha Ruibal, Sean Preci, Greta Castaneda, Armida Lizarraga, and Julie Russ for their instrumental roles in carrying out the study. The authors would also like to thank David J. Francis for his insights and, most importantly, thank the participating principals, teachers, and students.

## References

- Alexander, P.A., & Jetton, T.L. (2000). Learning from text: A multidimensional and developmental perspective. In R. Barr, M. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 3, pp. 285–310). New York: Longman.
- Anyon, J. (1980). Social class and the hidden curriculum of work. *Journal of Education*, 162(1), 67–92.
- Anyon, J. (1997). *Ghetto schooling: A political economy of urban educational reform*. New York: Teachers College Press.
- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice*, 20(1), 50–57. doi:10.1111/j.1540-5826.2005.00120.x
- August, D., & Hakuta, K.E. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.
- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth*. Mahwah, NJ: Erlbaum.
- Bailey, A.L. (2006). *Language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.

- Baumann, J.F., Edwards, E.C., Font, G., Tereshinski, C.A., Kame'enui, E.J., & Olejnik, S. (2002). Teaching morphemic and contextual analysis to fifth-grade students. *Reading Research Quarterly*, 37(2), 150–176.
- Baumann, J.F., Kame'enui, E.J., & Ash, G.E. (2003). Research on vocabulary instruction: Voltaire redux. In J. Flood, D. Lapp, J.R. Squire, & J.M. Jensen (Eds.), *Handbook of research on teaching the English language arts* (2nd ed., pp. 752–785). Mahwah, NJ: Erlbaum.
- 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(3), 251–271.
- Beck, I.L., McKeown, M.G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.
- Beech, J.R., & Keys, A. (1997). Reading, vocabulary and language preference in 7- to 8-year-old bilingual Asian children. *The British Journal of Educational Psychology*, 67(4), 405–414.
- Biancarosa, G., & Snow, C.E. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education. Retrieved December 3, 2007, from [www.carnegie.org/literacy/pdf/ReadingNext.pdf](http://www.carnegie.org/literacy/pdf/ReadingNext.pdf)
- Blachowicz, C.L.Z., & Fisher, P.J. (2007). Best practices in vocabulary instruction. In L. Gambrell, L. Morrow, & M. Pressley (Eds.), *Best practices in literacy instruction* (pp. 178–203). New York: Guilford.
- Bloom, H.S., Hill, C.J., Black, A.R., & Lipsey, M.W. (2008). *Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions*. New York: MDRC. Retrieved October 11, 2008, from [www.mdrc.org/publications/500/abstract.html](http://www.mdrc.org/publications/500/abstract.html)
- Bransford, J.D. (2004). Schema activation and schema acquisition: Comments on Richard C. Anderson's remarks. In R.B. Ruddell & N.J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 607–619). Newark, DE: International Reading Association.
- Bryk, A.S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage Foundation.
- Calderón, M. (2007). *Teaching reading to English language learners, grades 6–12: A framework for improving achievement in the content areas*. Thousand Oaks, CA: Corwin.
- California Department of Education. (1997). *English language arts content standards for California public schools: Kindergarten through grade twelve*. Sacramento: Author.
- Capps, R., Fix, M., Murray, J., Ost, J., Passel, J.S., & Herwatoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington, DC: The Urban Institute. Retrieved December 14, 2009, from [http://www.urban.org/UploadedPDF/311230\\_new\\_demography.pdf](http://www.urban.org/UploadedPDF/311230_new_demography.pdf)
- Carlisle, J.F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing: An Interdisciplinary Journal*, 12(3), 169–190. doi:10.1023/A:1008131926604
- Carlo, M.S., August, D., McLaughlin, B., Snow, C.E., Dressler, C., Lippman, D.N., et al. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, 39(2), 188–215. doi:10.1598/RRQ.39.2.3
- Carver, R.P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior*, 26(4), 413–437.
- Chall, J.S., & Jacobs, V.A. (2003). Poor children's fourth-grade slump. *American Educator*, 27(3), 14–15, 44.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Coyne, M.D., McCoach, D.B., Loftus, S., Zipoli, R., Jr., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching breadth vs. depth. *The Elementary School Journal*, 110(1), 1–18. doi:10.1086/598840
- Deshler, D., Palincsar, A., Biancarosa, G., & Nair, M. (2007). *Informed choices for struggling adolescent readers: A research-based guide to instructional programs and practices*. Newark, DE: International Reading Association.
- Donovan, M.S., Wigdor, A.K., & Snow, C.E. (2003). *Strategic education research partnership*. Washington, DC: The National Academies Press.
- Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first-and second-language learners. *Reading Research Quarterly*, 38(1), 78–103. doi:10.1598/RRQ.38.1.4
- Durkin, D. (1978). What classroom observations reveal about comprehension instruction. *Reading Research Quarterly*, 14(4), 481–533. doi:10.1598/RRQ.14.4.2
- Ellis, R. (1994). *The study of second language acquisition*. New York: Oxford University Press.
- Elmore, R.F. (2004). *School reform from the inside out: Policy, practice, and performance*. Cambridge, MA: Harvard University Press.
- Fang, Z., & Schleppegrell, M.J. (2008). *Reading in secondary content areas: A language-based pedagogy*. Ann Arbor: University of Michigan Press.
- Freebody, P., & Anderson, R.C. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, 15(3), 19–39.
- Frey, N. (2002). Literacy achievement in an urban middle-level professional development school: A learning community at work. *Reading Improvement*, 39(1), 3–13.
- Fukkink, R.G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. *Review of Educational Research*, 68(4), 450–469.
- Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, 11(36). Retrieved on October 7, 2003, from <http://epaa.asu.edu/epaa/v11n36/>
- Gersten, R., Dimino, J., & Jayanthi, M. (2007). Towards the development of a nuanced classroom observational system for studying comprehension and vocabulary instruction. In B. Taylor & J. Ysseldyke (Eds.), *Educational interventions for struggling readers, K-6* (pp. 381–425). New York: Teachers College Press
- 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). Newark, DE: International Reading Association; New York: Teachers College Press.
- Graves, M.F. (2006). *The vocabulary book: Learning and instruction*. New York: Teachers College Press.
- Graves, M.F., & Watts-Taffe, S.M. (2002). The place of word consciousness in a research-based vocabulary program. In A.E. Farstrup & S.J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 140–165). Newark, DE: International Reading Association.
- Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school: A follow-up. *The Journal of Special Education*, 34(2), 90–103. doi:10.1177/002246690003400204
- Guskey, T.R. (1989). Attitude and perceptual change in teachers. *International Journal of Educational Research*, 13(4), 439–453. doi:10.1016/0883-0355(89)90039-6
- Guskey, T.R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, 8(3), 381–391. doi:10.1080/135406002100000512
- Hakuta, K. (1998). Improving education for all children: Meeting the needs of language minority children. In D. Clark (Ed.), *Education*

- and the development of American youth (pp. 35–39). Washington, DC: The Aspen Institute.
- Hall, B. (2004). Literacy coaches: An evolving role. *Carnegie Reporter*, 3(1). Retrieved February 12, 2009, from [www.carnegie.org/reporter/09/literacy/index.html](http://www.carnegie.org/reporter/09/literacy/index.html)
- Hiebert, E.H. (2005). In pursuit of an effective, efficient vocabulary program. In E.H. Hiebert & M.L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 243–263). Mahwah, NJ: Erlbaum.
- Hutchinson, J.M., Whiteley, H.E., Smith, C.D., & Connors, L. (2003). The developmental progression of comprehension-related skills in children learning EAL. *Journal of Research in Reading*, 26(1), 19–32. doi:10.1111/1467-9817.261003
- Kieffer, M.J. (2008). Catching poverty, up or falling behind? Initial English proficiency, concentrated and the reading growth of language minority learners in the United States. *Journal of Educational Psychology*, 100(4), 851–868. doi:10.1037/0022-0663.100.4.851
- Kieffer, M.J., & Lesaux, N.K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing: An Interdisciplinary Journal*, 21(8), 783–804.
- Kintsch, W. (1994). Text comprehension, memory and learning. *American Psychologist*, 49(4), 294–303. doi:10.1037/0003-066X.49.4.294
- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37–62. doi:10.3102/01623737024001037
- Lee, V.E., Dedrick, R.F., & Smith, J.B. (1991). The effect of the social organization of schools on teachers' efficacy and satisfaction. *Sociology of Education*, 64(3), 190–208. doi:10.2307/2112851
- Leseman, P.P.M., & de Jong, P.F. (1998). Home literacy, opportunity, instruction, cooperation and social-emotional quality predicting early reading achievement. *Reading Research Quarterly*, 33(3), 294–318. doi:10.1598/RRQ.33.3.3
- Lively, T., August, D., Carlo, M., & Snow, C. (2003). *Vocabulary improvement program for English language learners and their classmates: 6th grade*. Baltimore, MD: Paul H. Brookes.
- Lublinter, S., & Smetana, L. (2005). The effects of comprehensive vocabulary on Title I students' metacognitive word-learning skills and reading comprehension. *Journal of Literacy Research*, 37(2), 163–200. doi:10.1207/s15548430jlr3702\_3
- Lutkus, A., Grigg, W., & Donahue, P. (2007). *The Nation's Report Card: Trial Urban District Assessment Reading 2007* (NCES 2008-455). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Manis, F.R., Lindsey, K.A., & Bailey, C.E. (2004). Development of reading in grades K–2 in Spanish-speaking English-language learners. *Learning Disabilities Research & Practice*, 19(4), 214–224.
- Marsh, J.A., McCombs, J.S., Lockwood, J.R., Martorell, F., Gershwin, D., Naftel, S., et al. (2008). *Supporting literacy across the sunshine state: A study of Florida middle school reading coaches*. Santa Monica, CA: RAND.
- Marzano, R.J., & Pickering, D.J. (2005). *Building academic vocabulary: Teacher's manual*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Maxwell, J.A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Thousand Oaks, CA: Sage.
- McKeown, M.G., & Beck, I.L. (2004). Direct and rich vocabulary instruction. In J.F. Baumann & E.J. Kame'enui (Eds.), *Vocabulary instruction: Research to practice* (pp. 13–27). New York: Guilford.
- McLaughlin, B. (1985). *Second-language acquisition in childhood, Volume 2: School age children*. Hillsdale, NJ: Erlbaum.
- McLaughlin, B., August, D., Snow, C., Carlo, M., Dressler, C., White, C., et al. (2000, April). *Vocabulary improvement in English language learners: An intervention study*. Symposium presented at the National Clearinghouse for English Language Acquisition, Washington, DC.
- McNamara, D.S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1–43.
- Moore, D.W., Bean, T.W., Birdyshaw, D., & Rycik, J.A. (1999). *Adolescent literacy: A position statement for the Commission on Adolescent Literacy of the International Reading Association*. Retrieved February 22, 2009, from [www.reading.org/downloads/positions/ps1036\\_adolescent.pdf](http://www.reading.org/downloads/positions/ps1036_adolescent.pdf)
- Nagy, W.E., Berninger, V.W., & Abbott, R.D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of Educational Psychology*, 98(1), 134–147. doi:10.1037/0022-0663.98.1.134
- Nagy, W.E., & Scott, J.A. (2000). Vocabulary processes. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 3, pp. 269–284). Mahwah, NJ: Erlbaum.
- Nair, M. (2007). *An analysis of the words appearing in middle school textbooks*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
- Nakamoto, J., Lindsey, K.A., & Manis, F.R. (2007). A longitudinal analysis of English language learners word decoding and reading comprehension. *Reading and Writing: An Interdisciplinary Journal*, 20(7), 691–719. doi:10.1007/s11145-006-9045-7
- National Center for Education Statistics. (2008). *NAEP 2008 trends in academic progress: Reading and mathematics*. Washington, DC: U.S. Department of Education.
- National Center for Education Statistics. (2009a). *The condition of education 2009*. Retrieved November 1, 2009, from <http://nces.ed.gov/programs/coe/list/index.asp>
- National Center for Education Statistics. (2009b). *Characteristics of public, private, and Bureau of Indian Education elementary and secondary schools in the United States: Results from the 2007–08 schools and staffing survey*. Retrieved October 30, 2009, from <http://nces.ed.gov/pubs2009/2009321.pdf>
- National Council of Teachers of English. (2006). *NCTE principles of adolescent literacy reform*. Urbana, IL: Author. Retrieved March 15, 2008, from [www.ncte.org/library/NCTEFiles/Resources/PolicyResearch/AdolLitPrinciples.pdf](http://www.ncte.org/library/NCTEFiles/Resources/PolicyResearch/AdolLitPrinciples.pdf)
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.
- Nelson, J.R., & Stage, S.A. (2007). Fostering the development of vocabulary knowledge and reading comprehension through contextually-based multiple meaning vocabulary instruction. *Education & Treatment of Children*, 30(1), 1–22. doi:10.1353/etc.2007.0003
- Pearson, P.D., Hiebert, E.H., & Kamil, M.L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, 42(2), 282–296.
- Peregoy, S.F., & Boyle, O.F. (2008). *Reading, writing, and learning in ESL: A resource book for teaching K–12 English learners* (5th ed.). Boston: Pearson Education.
- Pérez, E. (1981). Oral language competence improves reading skills of Mexican American third graders. *The Reading Teacher*, 35(1), 24–27.
- 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(2), 246–256. doi:10.1037/0022-0663.97.2.246
- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Rex, L.A., & Nelson, M.C. (2004). How teachers' professional identities position high-stakes test preparation in their

- classrooms. *Teachers College Record*, 106(6), 1288–1331. doi:10.1111/j.1467-9620.2004.00380.x
- Roser, N., & Juel, C. (1982). Effects of vocabulary instruction on reading comprehension. In J.A. Niles & L.A. Harris (Eds.), *New inquiries in reading research and instruction* (31st yearbook of the National Reading Conference, pp. 110–118). Rochester, NY: National Reading Conference.
- Ruddell, R.B., & Unrau, N.J. (2004). The role of responsive teaching in focusing reader intention and developing reader motivation. In R.B. Ruddell & N.J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 954–978). Newark, DE: International Reading Association.
- Saddler, B., & Graham, S. (2007). The relationship between writing knowledge and writing performance among more and less skilled writers. *Reading & Writing Quarterly*, 23(3), 231–247. doi:10.1080/10573560701277575
- Scarcella, R. (2003). *Academic English: A conceptual framework*. Los Angeles: University of California Language Minority Research Institute.
- Schleppegrell, M.J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Erlbaum.
- Schneider, B. (2004). Building a scientific community: The need for replication. *Teachers College Record*, 106(7), 1471–1483. doi:10.1111/j.1467-9620.2004.00386.x
- Scott, J.A., Jamieson-Noel, D., & Asselin, M. (2003). Vocabulary instruction throughout the day in twenty-three Canadian upper-elementary classrooms. *The Elementary School Journal*, 103(3), 269–283. doi:10.1086/499726
- Shanahan, T., & Beck, I.L. (2006). Effective literacy teaching for English-language learners. In D.L. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth* (pp. 415–488). Mahwah, NJ: Erlbaum.
- Slavin, R.E., Cheung, A., Groff, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best evidence synthesis. *Reading Research Quarterly*, 43(3), 290–322. doi:10.1598/RRQ.43.3.4
- Snow, C., Burns, M., & Griffin P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stahl, S.A. (1999). *Vocabulary development*. Cambridge, MA: Brookline Books.
- Stahl, S.A., & Fairbanks, M.M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, 56(1), 72–110.
- Stahl, S.A., & Nagy, W.E. (2006). *Teaching word meanings*. Mahwah, NJ: Erlbaum.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360–407. doi:10.1598/RRQ.21.4.1
- Sternberg, R.J. (1987). Most vocabulary is learned from context. In M.G. McKeown & M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89–105). Hillsdale, NJ: Erlbaum.
- Stoller, F., & Grabe, W. (1995). Implications for L2 vocabulary acquisition and instruction from L1 vocabulary research. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 24–45). Norwood, NJ: Ablex.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.
- Sumara, D.J. (2000). Researching complexity. *Journal of Literacy Research*, 32(2), 267–281.
- Swanborn, M.S.W., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69(3), 261–285.
- 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(2), 247–264. doi:10.1037/0022-0663.98.2.247
- Sweet, A.P., & Snow, C.E. (2004). *Rethinking reading comprehension*. New York: Guilford.
- Umbel, V.M., Pearson, B.Z., Fernandez, M.C., & Oller, D.K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development*, 63(4), 1012–1020. doi:10.2307/1131250
- Valdés, G. (2000). *Learning and not learning English: Latino students in American schools*. New York: Teachers College Press.
- Vaughn-Shavuo, F. (1990). *Using story grammar and language experience for improving recall and comprehension in the teaching of ESL to Spanish-dominant first-graders*. New York: Hofstra University.
- Verhoeven, L.T. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading*, 4(4), 313–330. doi:10.1207/S1532799XSSR0404\_4
- Wade, S.E., & Moje, E.B. (2000). The role of text in classroom learning. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 609–627). Mahwah, NJ: Erlbaum.
- Wagner, R.K., Muse, A.E., & Tannenbaum, K.R. (Eds.). (2007). *Vocabulary acquisition: Implications for reading comprehension*. New York: Guilford.
- Watts, S.M. (1995). Vocabulary instruction during reading lessons in six classrooms. *Journal of Reading Behavior*, 27(3), 399–424.
- Fillmore, L.W. (1982). Language minority students and school participation: What kind of English is needed? *Journal of Education*, 164(2), 143–156.
- Yoshikawa, H., Weisner, T.S., Kalil, A., & Way, N. (2008). Mixing qualitative and quantitative research in developmental science: Uses and methodological choices. *Developmental Psychology*, 44(2), 344–354. doi:10.1037/0012-1649.44.2.344

Submitted April 9, 2009

Final revision received November 7, 2009

Accepted November 13, 2009

**Nonie K. Lesaux** is Marie and Max Kargman Associate Professor in Human Development and Urban Education Advancement at the Harvard Graduate School of Education, Cambridge, Massachusetts, USA; e-mail lesauxno@gse.harvard.edu.

**Michael J. Kieffer** is Assistant Professor of Language and Education at Teachers College, Columbia University, New York, USA; e-mail mk3157@columbia.edu.

**S. Elisabeth Fallor** is a graduate student at the Harvard Graduate School of Education, Cambridge, Massachusetts, USA; e-mail sef416@mail.harvard.edu.

**Joan G. Kelley** is a research associate in the Language Diversity & Literacy Development research group at the Harvard Graduate School of Education, Cambridge, Massachusetts, USA; e-mail joangkelley@gmail.com.

# Sample One-Page Outline and Elaborated Instructional Model

**Target Words:** *affect, community, contribute, culture, ethnic, establish, resident, welfare*

## **Unit 1, Day 1. Introduction**

**Objective:** Students will be exposed to target words as they learn the central concepts and main idea of the text.

**What's the Point? (WTP):** Rather than simply memorizing vocabulary, students will be invested and engaged in learning concepts and acquiring background knowledge as they read the text, and then they will map the labels (words) onto these new ideas. Using the text provides a meaningful context for discussing the words throughout the unit.

**Materials:** world map, document camera or overhead projector, article

### **Lesson Outline**

1. Teacher—Preview article (10 minutes)

WTP? Students build background knowledge.

- Find Kibera on map
- Explain how a visitor improved lives in Kibera
- Ask students to share experiences about playing on a team

2. Teacher—Read aloud article (5 minutes)

WTP? Students need a meaningful and engaging context for learning the words.

3. Teacher—Introduce vocabulary words (5 minutes)

WTP? Students hear, see, and make note of the labels.

- Read each word that is highlighted in the article aloud
- Students repeat each word and then underline it
- Post list in classroom

4. Teacher—Introduce context clues (5 minutes)

WTP? Students learn to be better word-learners.

- Explain how context clues can be helpful (resident)
- Point out that the clue may not be in the same sentence (resident)
- Explain that sometimes context doesn't help very much (*establishing*)

## **1. Preview article**

**Say:** I have an article here that talks about kids who live in Africa in an area called Kibera. Can someone come up and help me find Kibera on the map? It's near Nairobi, a big African city. The article will tell us that Kibera is a poor area with a history of violence between different groups of people living there. The kids who grow up in Kibera don't feel like there is much hope for the future. A visitor to Kibera helped improve the town and got the kids excited about life by getting them together to play a game that a lot of you have played, I imagine. Soccer! Raise your hand if you have ever played soccer. Some of you might have played with friends (informally) and some of you might have been on soccer teams. Who has played on another kind of sports team? How do people get on teams?

**Do:** Lead class discussion on the various ways people become part of teams (trying out or signing up or paying to play).

## **2. Read article aloud**

## **3. Introduce vocabulary words**

**Say:** You saw the eight highlighted words in the article as I read. Those are the words we will be studying for the next couple of weeks; they are posted up front and will remain posted, so we can continually refer back to them in the coming days. Let's go over them. I'll read each word aloud, and I'd like you—as a class—to repeat after me. Saying words is a step to learning them, because you hear yourself speaking them versus just thinking them in your head. Another important reason to repeat

words aloud is because some words do not sound like what you would expect based on the spelling.

**Do:** Read first word aloud. Tell them to repeat after you.

**Say:** Let's add another step to this. After you repeat the word, use a pencil to underline it in the text.

**Do:** Read second word. Tell them to repeat after you. Remind them to underline. Repeat for next six words in the lesson.

#### **4. Introduce context clues**

**Say:** How many of you have heard of one or some of the highlighted words before? All readers—adults and kids—have the experience of reading something and not knowing what a word means. One strategy many adults and kids use is looking at the context clues; in fact, many people learn to do this automatically and may not be aware that they are using this strategy. Does anyone know what I mean by context clues? A context clue is a bit of information in a book or article that tells you something about a word you may not understand fully—in other words, it presents the context of the word, or what is around the word. Sometimes there are context clues in the same sentence, and sometimes the clues are in other places in the text. Let's start by looking at the word *resident*. Could someone please find the word *resident* in the text?

**Do:** Project article onto screen and ask a student to read aloud the sentence that contains the word *resident*. Underline the word *resident* on the screen.

**Say:** Does this sentence contain any hints about the meaning of the word *resident*? [Note: in this case it does not actually help that much.] Oftentimes the reader needs to go to the sentence before or the sentence after to figure out if there are any clues about the unknown word's meaning. Sometimes the clue is even further away from the word. Look at the sentences around the *resident* sentence and find the words that help the reader figure out the meaning of *resident* [*"people live in an area,"* from preceding sentence.]

**Do:** Underline the words *"people live in an area"* on the screen text.

**Say:** Now let's try another word from our list. Can someone find the sentence that uses the word *establishing*? [In the third paragraph from the end.] In this case, the context clues don't tell us what *establishing* means. It could mean having girls' soccer or starting a team or even talking about girls' soccer. *Establishing* actually means starting something. Context clues don't always tell you what the unfamiliar word means, but using context often does help so it is a good place to start when you see a new word in text.