

UC Irvine

UC Irvine Previously Published Works

Title

Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program

Permalink

<https://escholarship.org/uc/item/53f6m0c8>

Journal

Bilingualism: Language and Cognition, 15(3)

Authors

Lawrence, Joshua Fahey
Capotosto, Lauren
Branum-Martin, Lee
et al.

Publication Date

2012-07-01

License

<https://creativecommons.org/licenses/by/4.0/> 4.0

Peer reviewed

Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program*

JOSHUA F. LAWRENCE
Harvard University
LAUREN CAPOTOSTO
Harvard University
LEE BRANUM-MARTIN
University of Houston
CLAIRE WHITE
SERP Institute
CATHERINE E. SNOW
Harvard University

(Received: January 26, 2011; final revision received: June 24, 2011; accepted: July 1, 2011)

This longitudinal quasi-experimental study examines the effects of Word Generation, a middle-school vocabulary intervention, on the learning, maintenance, and consolidation of academic vocabulary for students from English-speaking homes, proficient English speakers from language-minority homes, and limited English-proficiency students. Using individual growth modeling, we found that students receiving Word Generation improved more on target word knowledge during the instructional period than students in comparison schools did, on average. We found an interaction between instruction and home-language status such that English-proficient students from language-minority homes improved more than English-proficient students from English-speaking homes. Limited English-proficiency students, however, did not realize gains equivalent to those of more proficient students from language-minority homes during the instructional period. We administered follow-up assessments in the fall after the instructional period ended and in the spring of the following year to determine how well students maintained and consolidated target academic words. Students in the intervention group maintained their relative improvements at both follow-up assessments.

Keywords: vocabulary, instruction, longitudinal analysis

Introduction

In 2008, approximately 10.9 million children aged 5–17 years in the United States spoke a language other than English in the home (Aud, Hussar, Planty, Snyder, Bianco,

Fox, Frohlich, Kemp & Drake, 2010). Compared with their native English-speaking peers, language-minority students have lower reading performance in English, on average (August & Shanahan, 2006). Although numerous factors account for this gap, researchers have pointed to differences in vocabulary knowledge as part of the explanation. Language-minority students have both less depth (Verhallen & Schoonen, 1993) and less breadth of vocabulary. Although the causal link between reading comprehension and vocabulary size has not been proved (National Institute of Child Health and Human Development, 2000), a high proportion of unknown words in a given text can disrupt comprehension of it (Carver, 1994). Just as students from English-speaking homes encounter new reading difficulties in the upper grades when vocabulary demands in texts increase (Chall & Jacobs, 2003) and the words encountered become more abstract and academic (Scarcella, 2003), so, too, do language-minority learners, perhaps to an even greater degree.

Some research suggests that language-minority students in the middle grades may benefit from explicit vocabulary instruction that involves multiple exposures to target words in diverse contexts (Carlo, August,

* The SERP-BPS field site and thus the original planning for Word Generation were supported by grants to the Strategic Education Research Partnership (SERP) from the Spencer Foundation and the William and Flora Hewlett Foundation; further development and evaluation of Word Generation were supported by a Senior Urban Education Fellowship awarded to Catherine Snow by the Council of Great City Schools. Joshua Lawrence was supported by funds awarded to Catherine Snow by the Spencer Foundation and the Carnegie Corporation of New York. We also acknowledge the funding to SERP from the Lowenstein Foundation, to develop professional development opportunities through www.wordgeneration.org. The first author was supported by Grant Number R305A090555, Word Generation: An Efficacy Trial from the Institute of Educational Sciences (IES), US Department of Education (USDE) during the preparation of this paper. Additional support was received from Grant Number R305A050056, National Research and Development Center for English Language Learners. The contents do not necessarily represent the positions or policies of IES or USDE and readers should not assume endorsement by the federal government for any of the positions or statements expressed herein. Our thanks to the anonymous reviewers of this journal for their insightful comments.

Address for correspondence:

Joshua F. Lawrence, Department of Education, University of California, Irvine, 3200 Education Building, Irvine, CA 92697-5500, USA
jflawren@uci.edu

McLaughlin, Snow, Dressler, Lippman, Livey & White, 2004; Proctor, Dalton, Uccelli, Biancarosa, Mo, Snow & Neugebauer, 2009/2011; Snow, Lawrence & White, 2009; Vaughn, Martinez, Linan-Thompson, Reutebuch, Carlson & Francis, 2009). The current study aims to build upon this work. It is based upon an unmatched quasi-experiment conducted in close cooperation with Boston Public Schools that investigates the effects of Word Generation (WG), a cross-content academic language intervention program, on the vocabulary performance of sixth- to eighth-grade students. The program was created during the year before this quasi-experiment was conducted by some authors of this paper in close collaboration with Boston teachers. The program teaches five all-purpose academic words each week. Beck, McKeown and Kucan (2002) suggest a rough heuristic for categorizing words as those that most school-aged children will know (tier-one words), those that students are only likely to encounter in texts for one content area (tier-three words), and others that are not well known, but might appear in any number of academic content areas (tier-two words). One source for identifying all-purpose academic words is The Academic Word List, which was developed by analyzing a range of adult academic texts to identify words that were used in multiple academic contexts across genres (Coxhead, 2000). Examples include *distribute*, *conclusion*, *proceed*, *logical*, *obtain*, *acquire*, *retain*, *exclude*, *attribute*, *assume*, *capacity*, *enable*, *perspective*, *relevant*, *perceive*, *component*, *restrict*, *generate*, *distinct*, *assess*, *alter*, *amend*, and *contrast*. We used the Coxhead list and other sources (Lawrence, White & Snow, 2010) to identify appropriate all-purpose academic words.

The target words for each week of instruction are embedded in a high interest passage about a controversial topic that is read by students in English classes on Monday. On each of the next three weekdays one of the content teachers delivers a 15-minute lesson that is related to the overarching topic but presents the target words in content-specific contexts. For instance, on Tuesday the social studies teacher may facilitate a debate about if pet rentals should be legal, highly regulated or unregulated. Because Tuesday would be the second day that students have thought about this topic and encountered the academic language, the teacher will have less scaffolding to do to support their use of the academic language. On Wednesday, the math teacher may have students answer a math word problem that presents data based on the number of hours that customers rent pets for and then ask them to determine the median number of rental hours. On Thursday, the science teacher introduces fictitious experimental data about dog happiness and asks students to draw inferences. On Friday, the English teacher asks students to “take a stand” by responding to a persuasive writing prompts about whether the benefits of renting a pet outweigh the potential harm it causes animals.

In the first study that resulted from this work (Snow et al., 2009), we found that students in Boston middle schools implementing Word Generation had greater one-time vocabulary gains than students in comparison schools, such that students in the Word Generation program learned approximately the number of words that differentiated eighth from sixth graders on the pretest – in other words, program participation resulted in gains equivalent to two years of incidental word learning. Furthermore, the language-minority students in the Word Generation, but not the comparison, schools showed greater gains than the English-only students. That study provides mean pretest and posttest scores for all the items in the first year of the study, and more details about program implementation. The current longitudinal study extends this work by following up on participating students after summer vacation and then one full year after instructional sessions. Thus the current paper examines not only how well students from language-minority homes learn academic vocabulary, but also how well they maintain vocabulary knowledge in their second language. Furthermore, the current study extends our initial study by examining not only home-language status but also language proficiency as a predictor of vocabulary learning and maintenance.

Background and context

Children come to understand the multiple meanings and uses of words through repeated encounters with them (Fukkink & de Glopper, 1998; Nagy & Scott, 2000). Not surprisingly then, children’s knowledge of high-frequency words is unlikely to decay, and may even expand, if they are in settings where they continue to encounter these words frequently.

Guided by this knowledge, a few studies have examined the impact of vocabulary interventions that promote many exposures to words for English language learners in the middle grades. These studies commonly examined the impact of instruction of target words in rich contexts, but differed in their program features (see Table 1). For instance, Word Generation (Snow et al., 2009) is a cross-content vocabulary program that teaches general purpose academic vocabulary words in language arts, mathematics, science, and social studies classrooms. In contrast, Quality English and Science Teaching (QuEST) (August, Branum-Martin, Cardenas-Hagan & Francis, 2009) promotes language development in the science classroom, while a program developed by Vaughn et al. (2009) provides direct instruction of academic vocabulary in social studies. The programs also differ in their target students. Some programs, such as the Vocabulary Improvement Program (VIP) (Carlo et al., 2004), QuEST (August et al., 2009), and Language Workshop (Townsend & Collins, 2009) were explicitly designed for use with language-minority students.

Table 1. *Characteristics of vocabulary studies that include English language learners (ELLs) in the middle grades.*

| Program/authors | Program description | Program features specific to ELLs | Language by treatment interaction |
|---|---|--|---|
| Vocabulary Improvement Program (VIP) (Carlo et al., 2004) | With a focus on both target word instruction and word-learning strategies, this program presents target words in engaging texts to ensure recurrent exposure. | Instruction in cognates and text previews in Spanish. | Yes; ELLs in treatment schools improved more than ELH students on polysemy task. |
| Word Generation (Snow et al., 2009) | This cross-content vocabulary program provides direct instruction in general purpose academic vocabulary words in language arts, mathematics, science, and social studies. | None. | Yes; LM students showed greater gains than ELH students in treatment, but not comparison schools. |
| Vaughn et al. (2009) | Students in social studies classrooms receive direct instruction of academic vocabulary, encounter target words in texts and videos, and participate in structured paired groupings. | Students use graphic organizers and writing to learn relationships between Spanish and English words. | No. |
| Quality English and Science Teaching (QuEST; August et al., 2009) | This program promotes science knowledge through hands-on experimentation and language development through explicit instruction of general academic and science vocabulary. | Instruction uses visual images and Spanish translations to support ELLs. | No. |
| Improving Comprehension Online (Proctor et al., 2009/2011) | Students read short digital texts that include supports aligned with principles of Universal Design for Learning (Rose & Meyer, 2002), including audio readings, multimedia glossaries, and illustrations to support comprehension. | Supports include Spanish translations, a human readings of text in English and Spanish, and bilingual pedagogical coaches to provide assistance. | No. |
| Language Workshop (Townsend & Collins, 2009) | This after-school intervention incorporates strategies for identifying and using cognates. | Emphasis on Spanish cognates. | Not applicable; all Spanish–English speakers. |

Accordingly, these programs offer instructional features designed specifically for the needs of English language learners, including the use of graphic organizers to learn relationships between English and Spanish words (Vaughn et al., 2009), text previews in Spanish (Carlo et al., 2004), Spanish translations (QuEST; August et al., 2009), and instruction in Spanish cognates (August et al., 2009; Carlo et al., 2004; Townsend & Collins, 2009). In contrast, Word Generation was designed for a general student population and has been used with students from both English-only and language-minority homes.

English language learners participating in vocabulary programs have outperformed their comparison group peers on curriculum-based measures of vocabulary (August et al., 2009; Carlo et al., 2004; Proctor et al., 2009/2011; Snow et al., 2009; Vaughn et al., 2009), science (August et al., 2009), and comprehension (Vaughn et al., 2009). These studies differed, however,

in whether they found varying effects for students of different language groups. For instance, ELLs participating in VIP improved as much as English-only students on word mastery, word association, and cloze tasks, but outperformed English-only students on a polysemy task (Carlo et al., 2004). Similarly, Snow et al. (2009) found that students from language-minority homes showed greater growth on a researcher-designed vocabulary measure than English-only students in Word Generation treatment schools, but not comparison schools. In contrast, studies of QuEST (August et al., 2009), Improving Comprehension Online (Proctor et al., 2009/2011), and Vaughn et al.'s (2009) intervention showed no difference in effects between English-only and English language learners.

While these studies examined only immediate impacts and used primarily curriculum-based measures, they suggest that explicit vocabulary instruction may help improve

the word knowledge of English language learners. At the same time, they highlight a need for further research. First, studies that have tested for interaction effects between treatment and language proficiency found a range of potential effects with some finding no difference between the effects for English-proficient and ELL students (e.g., August et al., 2009; Proctor et al., 2009/2011) and others finding that students from language-minority backgrounds benefited more from treatment (e.g., Carlo et al., 2004; Snow et al., 2009). Identifying interventions from which all students benefit but ELLs gain even more is an important step toward improving literacy broadly and closing the achievement gap between English-proficient and ELLs specifically. Second, studies that have tested for differential effects have also only examined the impact of instruction for two broad groups of students – English-proficient and language-minority learners. Although such distinctions are common, the language-minority population is remarkably heterogeneous, composed of individuals who speak a language other than English in the home, those with limited English proficiency, those proficient in two or more languages, and English dominant students (August & Shanahan, 2006; Kieffer, 2008). Given these differences, it is crucial that we move beyond a dichotomous construction of language status when examining the effects of vocabulary interventions, as diverse groups may experience the same intervention differently.

Finally, no vocabulary study of English language learners in middle schools has examined the long-term impact of instruction. Such information is important, as students from low-income families tend not to improve in vocabulary knowledge during summer months at the rates their wealthier peers do, and many students actually regress in their word knowledge during the summer (Alexander, Entwisle & Olson, 2001, 2007; Entwisle, Alexander & Olson, 1997; Heyns, 1978). Students who come from homes where a language other than English is spoken are even less likely to encounter academic English words during summer months, a plausible explanation for why in one study these students experienced a greater summer setback than their peers from English-speaking homes even controlling for socioeconomic status (Lawrence, *in press*).

Foreign language research further highlights the importance of examining long-term impacts of vocabulary instruction (for a review see Bardovi-Harlig & Stringer, 2010). For example, de la Fuente (2006) examined the long-term effectiveness of second language vocabulary instruction on Spanish-word learning by native English speakers in a non-immersion setting. De la Fuente found no differences in vocabulary knowledge of the students who received enhanced instruction and traditional instruction immediately after instruction; however, students in the intervention group maintained target

vocabulary knowledge better so at the delayed posttest there were differences between the vocabulary skills of treatment and comparison students. Similarly, comparing the success of Chinese-speaking students' success in learning new English words from textual encounters with and without instructional support, Min (2008) found students in both conditions improved in their knowledge of target words, but those with instructional support performed better than those without. In a follow-up posttest both groups experienced significant vocabulary knowledge loss resulting in a reduced but still significant advantage for the group that received instructional support. Long-term studies are needed to determine whether similar patterns of attrition hold for middle school students participating in a vocabulary intervention.

The goal of the present study is to understand the long- and short-term effects of participation in the Word Generation program for three groups of students: proficient English speakers from English-language homes (ELH), proficient English speakers from language-minority homes (LMH), and limited English-proficient (LEP) students (there are small numbers of LEP students whose parents reported speaking English at home, and although they were included in this analysis we do not highlight this profile of student in our results as there are so few of them). In addition to pre- and immediate posttest data on words taught during the program, we tested eleven words again in fall and spring of the following academic year. We intend to determine both if participation in Word Generation benefits all students irrespective of home language status and proficiency, and if all groups of students maintain knowledge of target words relative to comparison students. Thus, our research questions (RQs) are:

RQ1. How did English speaking students from English-language homes (ELH) who participated in the Word Generation program learn, maintain, and consolidate words compared with similar students attending comparison schools?

RQ2. How did English-proficient students from language-minority homes (LMH) who participated in the Word Generation program learn, maintain, and consolidate words compared with similar students attending comparison schools?

RQ3. How did students with limited English proficiency (LEP) from language-minority homes who participated in the Word Generation program learn, maintain, and consolidate words compared with similar students attending comparison schools?

Methods

This study is based on data collected from an unmatched quasi-experiment conducted to determine the efficacy of

the Word Generation program. During the first year of this quasiexperiment, pre- and posttest data were collected from five treatment schools and four comparison schools. Students in the Word Generation schools received explicit vocabulary instruction for approximately fifteen minutes per day, as described above. Students in comparison schools received “business as usual” instruction where we observed different relative emphasis on content-specific vocabulary instruction in different classes but consistently limited instruction of high leverage cross-content vocabulary.

District setting

The study was conducted in the Boston Public Schools (BPS) through the Strategic Education Research Partnership (SERP), a nonprofit organization that aims to support sustained collaboration between educational researchers and public school districts. The Word Generation program was created in response to the district’s need for improved materials to support student literacy in middle schools. One year before the start of this study, the Word Generation program had been piloted in two Boston middle schools and redesigned based on feedback solicited from pilot teachers. To better understand the impact of the program, SERP and BPS arranged to conduct a quasi-experiment, with program implementation in five schools and comparison data collected from four others. The schools that implemented the Word Generation program were volunteered by their principal to do so, the schools that did not were nominated by the district leadership. School leaders accepted a small financial incentive to the school for its cooperation. These differential selection criteria probably contributed to the fact that at baseline treatment and comparison schools were not well matched.

Boston has been recognized as a strong urban school district; it received the Broad Foundation prize in 2006, and is one of the highest performing urban districts in national measures of literacy (Lutkus, Rampey & Donahue, 2005). Like most urban districts in the United States, in 2007 Boston served many students from low-income families (74.3%), students whose first language was not English (38.1%) and students designated as limited English proficiency (LEP, 18.9%). District average student-level demographic indicators (available from the Massachusetts Department of Elementary and Secondary Education) are crucial in determining school and district performance levels according to federal assessment regulations (U.S. Department of Education, 2001). Definitions of these language and demographic categories are policy-driven rather than based directly on test scores. LEP designation indicates that students are receiving English development support from the school at the time of designation, or have in the previous two

years. The removal of the LEP designation is based on a number of factors including state achievement tests, teacher recommendations, and grades. Although there are district guidelines for this designation and re-designation process, there is considerable discretion in how it is completed by schools.

Procedure

In the first year of the quasi-experiment, students in the treatment schools received instruction on 120 high leverage academic words. To assess the impact of the study, students in both the treatment and comparison schools completed a pre- and posttest on their knowledge of 40 of the instructed target words (in the fall of 2007 and the spring of 2008). The third (fall 2008) and fourth (spring 2009) waves of data were collected primarily to assess the effectiveness of the second year of the Word Generation quasi-experiment. On each of these occasions students completed 50 multiple-choice items, the majority of which tested words instructed during the second year. However, 11 items taken from the previous year’s test were embedded in these assessments in order to conduct these longitudinal analyses. To construct a longitudinally consistent measure and maximize the amount of information from these 11 items tested four times, we used an item response theory (IRT) approach. First, we fit a single-factor model to the 11 items in each wave to test the hypothesis that the 11 items were reasonable indicators of a single factor of vocabulary knowledge. Then, we used the item parameters from wave one to produce scaled scores for each of the subsequent waves. Details on this scaling process are given in the Results section.

Longitudinal analytical methods allow the flexible use of data (Singer & Willett, 2003). This flexibility allowed us to include all students who contributed at least one wave of data during the first year (fall 2007 – spring 2008) in our analysis, although we did not include students who contributed data only during the third (fall 2008) or fourth (spring 2009) waves because we could not be sure that these students had received instruction on the target words and we were worried about the high mobility rates of our LEP students. This process resulted in no cases being dropped for the first two waves of data but the exclusion of students who entered the study during the second year. This process also allowed us to use data from eighth-grade students to help specify initial status and instructional impact, even if they did not contribute data to the follow-up analysis because they graduated from the participating schools and moved to high school.

The available data for this study based on these inclusion criteria are presented in Table 2. The first data column of this table shows the number of students who contributed data at each wave of collection. Scanning down this column demonstrates an attrition of the available

Table 2. Number of students who contributed to each wave of data collection by home language and English proficiency status in treatment and comparison schools.

| | | Number of students who contributed data at each wave | | | | |
|----------------------------|-------------|--|---------|-----|---------|-----|
| | | Total | LMH | | EH | |
| | | | Non-LEP | LEP | Non-LEP | LEP |
| Comparison school students | Fall 2007 | 525 | 164 | 33 | 323 | 5 |
| | Spring 2008 | 405 | 137 | 30 | 233 | 5 |
| | Fall 2008 | 204 | 54 | 27 | 121 | 2 |
| | Spring 2009 | 257 | 83 | 22 | 149 | 3 |
| Treatment school students | Fall 2007 | 1140 | 329 | 82 | 719 | 10 |
| | Spring 2008 | 1179 | 324 | 84 | 758 | 13 |
| | Fall 2008 | 757 | 185 | 62 | 500 | 10 |
| | Spring 2009 | 680 | 175 | 55 | 445 | 5 |

sample due, in part, to the oldest students graduating at the end of the first year, as well as student movement within and beyond the district. Looking across rows in Table 2 reveals that, while the parents of most LEP students asked to communicate with the school in a language other than English, some LEP students' parents were on record as wishing to communicate with the school in English. For instance, the top row of Table 2 shows that of the 197 language-minority students contributing data from the comparison schools at the first wave, 33 (around 18%) were identified as LEP by the district. Of the 328 students from English-speaking homes, five (around 1.5%) were identified as LEP. In the current analysis we include both home-language status and English proficiency level as independent variables and model results for each of the four subcategories that result. Although LEP students whose parents or guardians speak English at home no doubt constitute an intriguing subsample likely to have experienced family reunification, adoption, or other challenging experiences (Suárez-Orozco, Suárez-Orozco & Todorova, 2008), we have so few of these students that we do not differentiate them in our findings section.

Measures

Vocabulary

The 11 items that make up the vocabulary score in the current study are a subsample of words instructed and tested during the first year of the quasi-experiment in Boston that were subsequently embedded in the pre- and posttests during the following year. The target words in the subsample were: *acquire*, *contrast*, *disproportionate*, *enables*, *enforced*, *generate*, *incentives*, *interact*, *obtain*, *paralyzed*, and *relevant*. Each of the target words is taken from a list of academic words (Coxhead, 2000). Each of the 11 items was scored correct/incorrect and these were analyzed with an item response theory (IRT) model

Table 3. Fit statistics for categorical confirmatory factor analysis (CFA) models for each wave.

| Wave | Chi-square (df) | CFI | RMSEA | WRMR |
|------|-----------------|-------|-------|------|
| 1 | 111.1 (41) | 0.941 | 0.031 | 1.21 |
| 2 | 100.4 (42) | 0.962 | 0.028 | 1.12 |
| 3 | 125.1 (42) | 0.958 | 0.027 | 1.25 |
| 4 | 67.1 (42) | 0.99 | 0.016 | 0.9 |

CFI = comparative fit index; RMSEA = root mean square error of approximation; WRMR = weighted root mean square residual

Note: All models fit with robust weighted least squares estimation (WLSMV; Muthén & Muthén, 2007).

which formed a time-varying level-1 outcome VOCAB. The IRT scaled score was produced by fitting a single factor confirmatory factor analysis model to the eleven items separately for each wave, using Mplus 5, with robust weighted least squares estimation for dichotomous data (WLSMV; Muthén & Muthén, 2007). The model fit reasonably well in all four waves, as shown in Table 3. While there was some degree of misfit in the first wave (CFI = .94), the root mean square error of approximation was quite acceptable for all waves (RMSEA \leq .03). The coefficient alpha for each of the respective waves was 0.88, 0.86, 0.86 and 0.87. The item parameters (loadings and thresholds) from the first wave were then used to score the following three waves, thereby estimating a factor score on the metric of the first wave, with factor means and variances free to differ over time. In this way, the vocabulary scores for each wave were estimated on a single, consistent metric, relative to the first wave.

Wave

WAVE is a level-1 variable indicating wave of data collection (0 through 3).

Instruction

INSTRUCTION is a time-varying individual (level-1) variable that indicates how many instructional encounters students have had with target words. Students in Word Generation schools were instructed on these target words during the first but not second year, so the variable for those students is coded as follows: wave 0 = 0, wave 1 = 1, wave 2 = 1, wave 3 = 1. Comparison-school students were not explicitly instructed on these words, so INSTRUCTION was coded as 0 for them at each wave.

Summer

SUMMER indicates how many summers students had experienced since the start of the study (wave 0 = 0, wave 1 = 0, wave 2 = 1, wave 3 = 1); it is a time-varying continuous individual (level-1) variable.

Attends a Word Generation School

The measure WG_SCHOOL indicates if students attended a Word Generation school (WG_SCHOOL = 1) or a comparison school (WG_SCHOOL = 0). It is a level-2 variable.

Language-minority home

Language-minority home (LMH) is a level-2 variable indicating if a student's parent has requested to communicate with the school district in a language other than English (LMH = 1) or not (LMH = 0).

Limited English proficiency (LEP)

Limited English proficiency (LEP) is a level-2 variable indicating if a student had been admitted into the school system during the last two school years and was therefore eligible for bilingual support by the school during the first year of the study (LEP = 1) or not (LEP = 0).

Grade-level cohort

Grade level was provided by the school district and used to create two variables. GRADE7 describes if the student was in seventh grade (GRADE7 = 1) or not (GRADE7 = 0). GRADE8 describes if the student was in eighth grade (GRADE8 = 1) or not (GRADE8 = 0). This variable allows estimation of mean differences by grade.

Analysis

We used the multilevel model for change (Singer & Willett, 2003) to address each of the research questions. Power analysis revealed that although we expected treatment effect at the school level, we did not have sufficient schools in the study to analyze differences in growth at the school level and analyzed these data with a two-level rather than a three-level approach. Due to the limited number of waves of data available we assumed that

growth was linear, but included a parameter for summer setback. Level-2 variance (among students) in the rate-of-change parameter was negligible in all fitted models so it was fixed to zero. All models that were considered in determining the final fitted model were based on the exploration of a level-1, level-2 model with the following specifications:

Level-1 (outcomes in four waves across two years):

$$\widehat{VOCAB} = \pi_{0i} + \pi_{1i}WAVE_{ij} + \pi_{2i}INSTRUCTION_{ij} + \pi_{3i}SUMMER_{ij} + \varepsilon_{ij}$$

Level 2 (student level):

$$\begin{aligned}\pi_{0i} &= \gamma_{00} + \gamma_{01}GRADE7_i + \gamma_{02}GRADE8_i \\ &\quad + \gamma_{03}WG_SCHOOL_i + \gamma_{04}LMH_i + \gamma_{05}LEP_i + \zeta_{0i} \\ \pi_{1i} &= \gamma_{10} + \gamma_{11}GRADE7_i + \gamma_{12}GRADE8_i \\ &\quad + \gamma_{13}WG_SCHOOL_i + \gamma_{14}LMH_i + \gamma_{15}LEP_i \\ &\quad + \gamma_{16}LMH_iWG_SCHOOL_i \\ &\quad + \gamma_{17}LEP_iWG_SCHOOL_i \\ \pi_{2i} &= \gamma_{20} + \gamma_{21}GRADE7_i + \gamma_{22}GRADE8_i + \gamma_{23}LMH_i \\ &\quad + \gamma_{24}LEP_i \\ \pi_{3i} &= \gamma_{30} + \gamma_{31}WG_SCHOOL_i + \gamma_{32}LMH_i + \gamma_{33}LEP_i \\ &\quad + \gamma_{34}LMH_iWG_SCHOOL_i \\ &\quad + \gamma_{35}LEP_iWG_SCHOOL_i\end{aligned}$$

where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

This model allows us to use all waves of data from each student to create a model of vocabulary growth that examines potential improvement during the instructional period controlling for expected growth across the two years of the study and possible vocabulary setback during the summer months. Traditional methods allow analysis of changes between two waves of data collection but cannot model sophisticated growth trajectories across several waves of data such as is required to answer our research questions. The first research question, which asks about how ELH students in the WG program learned, maintained and consolidated words compared with ELH student in the comparison group, will be answered with reference to γ_{20} , $\gamma_{31}WG_SCHOOL_i$ and $\gamma_{13}WG_SCHOOL_i$ respectively. The second research question, which asks how English-proficient students from language-minority homes in the WG program learned, maintained and consolidated words relative to LMH students in the comparison schools will be answered by inspecting the parameters associated with the main effects of home-language status on the slope and summer setback ($\gamma_{14}LMH_i$ and $\gamma_{32}LMH_i$) and interaction between

the parameter associated with WG participation and home-language status ($\gamma_{23}LMH_i$, $\gamma_{34}WG_SCHOOL_i$, $\gamma_{16}LMH_iWG_SCHOOL_i$). Research question three asks about how LEP students who participated in the Word Generation program learned, maintained, and consolidated vocabulary knowledge compared to LEP student in comparison schools. Almost all the LEP students are from language-minority homes, so this question will be answered with reference to the parameters examined for RQ2. However, we also need to examine estimates of $\gamma_{15}LEP_i$, $\gamma_{24}LEP_i$, $\gamma_{33}LEP_i$, $\gamma_{35}LEP_iWG_SCHOOL_i$ and $\gamma_{17}LEP_iWG_SCHOOL_i$ to determine the additional impact of LEP status on word learning growth, and if LEP status interacts with participation in the WG program.

Results

The first data column of Table 4 provides the average scaled vocabulary achievement level for each treatment and comparison school at baseline (fall 2007). The second column of Table 4 presents the same statistics for the immediate posttest collected during spring 2008. Data columns three and four of Table 4 present scaled data from the third (fall 2008) and fourth (spring 2009) waves of data collection. The raw scores at each wave are presented on the right-hand columns. Scanning left to right across the first four rows suggests that students in both treatment and comparison schools tended to improve in word knowledge across successive waves of data collection except for a decline during summer months. This view also shows that there was attrition of the sample because students who started the study in eighth grade graduated to high schools. These descriptive data also suggest average improvement in treatment schools was larger ($M_{\text{wave1}} - M_{\text{wave4}} = 0.52$, scaled score) than improvement in the comparison schools ($M_{\text{wave1}} - M_{\text{wave4}} = 0.34$, scaled score). This table also demonstrates that some schools did not contribute data at each wave of data collection. These omissions were due to district-level reorganization and school closing in one case and logistical oversight in another.

Table 5 presents vocabulary data from comparison school and treatment school students across the four waves of data by home language and English proficiency status. This table suggests that English-proficient language-minority students began the study with slightly stronger vocabulary knowledge than English-proficient students from English-speaking homes. Examining baseline (fall 2007) scores demonstrates that comparison school students (top half of the first column) in all home-language and language-proficiency categories began the study with better vocabulary knowledge than their treatment peers on average (bottom half of the first column). Differences between English proficient and LEP students were pronounced at the baseline and

throughout the four waves of data collection for both treatment and comparison school students. Although these cross-sectional descriptive data provide a preliminary understanding of differences among subgroups, they do not account for the individual growth trajectories of students in the sample nor do they allow us to answer sophisticated questions about the impact of treatment by language proficiency level and home-language status across the four waves of data collected controlling for summer setback. To answer these research questions we must use individual growth modeling methods.

Table 6 presents the results of fitting a series of multilevel models for change predicting VOCAB across four waves of data. In the final fitted model, estimates are provided for several parameters that describe baseline population average vocabulary. The parameter estimate associated with the eighth-grade cohort was significant ($\gamma_{02}GRADE8_i = 0.297$, $p < .001$), which indicates that at the baseline, students in eighth grade scored higher than their sixth-grade peers on the vocabulary assessment, although sixth and seventh grade scores were indistinguishable at baseline. The parameter estimate for the term associated with being in eighth grade also interacted with instruction: eighth-grade students did not benefit as much from instruction ($\gamma_{22}GRADE8_i = -0.136$, $p < .01$). In fact, a general linear hypothesis (GLH; for more information see Singer & Willett, 2003, pp. 123–126) test shows that after accounting for this interaction term, there was no effect of treatment for eighth-grade students from English-speaking homes ($\chi^2 = 0.52$, *ns*). There were no differences in the benefit that sixth or seventh graders benefited from instruction.

The parameter estimate associated with treatment group ($\gamma_{03}WG_SCHOOL_i = -0.309$, $p < .001$) indicates that there were significant differences in average student performance between the treatment and comparison schools at the start of the study. English-proficient students from language-minority homes started the study with better vocabulary scores than students from English-speaking homes on average ($\gamma_{04}LMH_i = 0.138$, $p < .001$), but LEP students started the study at a significant disadvantage compared to their more English-proficient peers ($\gamma_{05}LEP_i = -0.528$, $p < .001$). These differences can be seen in the fall 2007 scores in the prototypical plots presented in Figure 1. The top two trajectories represent the average scores of language-minority (thick dashed line with markers) and English-home (thick dashed line) students in the comparison schools. The next two trajectories represent the population average scores of language-minority (thick solid line with markers) and English (thick solid line) homes in the treatment schools. The fifth line down represents the population average scores of LMH limited English-proficiency students in the comparison schools (thin dashed line). The bottom line represents the scores of LMH limited English-proficiency

Table 4. Average vocabulary scores on all eleven longitudinal items by wave by school.

| School | | Scaled | | | | Raw | | | |
|------------|------|--------------------|-------------|----------------|-------------|--------------------|-------------|----------------|-------------|
| | | Instructional year | | Follow-up year | | Instructional year | | Follow-up year | |
| | | Fall 2007 | Spring 2008 | Fall 2008 | Spring 2009 | Fall 2007 | Spring 2008 | Fall 2008 | Spring 2009 |
| Treatment | | | | | | | | | |
| Reilly | Mean | −0.088 | 0.473 | 0.108 | 0.442 | 4.666 | 6.099 | 5.682 | 6.719 |
| | SD | (0.728) | (0.793) | (0.772) | (0.826) | (2.122) | (2.172) | (2.306) | (2.460) |
| | N | 329 | 382 | 223 | 210 | 329 | 382 | 223 | 210 |
| Mercer | Mean | −0.047 | 0.445 | 0.098 | 0.487 | 4.835 | 5.841 | 5.674 | 6.562 |
| | SD | (0.752) | (0.859) | (0.835) | (0.795) | (2.194) | (2.443) | (2.448) | (2.271) |
| | N | 468 | 391 | 279 | 267 | 468 | 391 | 279 | 267 |
| Westfield | Mean | −0.215 | 0.195 | −0.193 | 0.262 | 4.254 | 5.116 | 4.679 | 5.971 |
| | SD | (0.672) | (0.786) | (0.832) | (0.864) | (2.069) | (2.227) | (2.422) | (2.671) |
| | N | 114 | 155 | 109 | 68 | 114 | 155 | 109 | 68 |
| Mystic | Mean | −0.017 | 0.559 | 0.150 | 0.491 | 4.883 | 6.134 | 5.687 | 6.670 |
| | SD | (0.705) | (0.803) | (0.712) | (0.873) | (2.019) | (2.192) | (2.044) | (2.478) |
| | N | 137 | 149 | 99 | 97 | 137 | 149 | 99 | 97 |
| Occidental | Mean | −0.305 | 0.214 | −0.355 | 0.132 | 4.087 | 5.431 | 4.064 | 5.684 |
| | SD | (0.672) | (0.890) | (0.639) | (0.803) | (1.948) | (2.507) | (2.151) | (2.395) |
| | N | 92 | 102 | 47 | 38 | 92 | 102 | 47 | 38 |
| Average | Mean | −0.093 | 0.416 | 0.038 | 0.431 | 4.674 | 5.831 | 5.435 | 6.518 |
| | SD | (0.765) | (0.872) | (0.780) | (0.856) | (2.132) | (2.326) | (2.382) | (2.419) |
| | N | 1140 | 1179 | 757 | 680 | 1140 | 1179 | 757 | 680 |
| Comparison | | | | | | | | | |
| Walters | Mean | 0.227 | n.a. | n.a. | n.a. | 5.696 | n.a. | n.a. | n.a. |
| | SD | (0.687) | n.a. | n.a. | n.a. | (2.031) | n.a. | n.a. | n.a. |
| | N | 92 | 0 | 0 | 0 | 92 | 0 | 0 | 0 |
| Garfield | Mean | 0.096 | 0.396 | n.a. | 0.308 | 5.375 | 5.754 | n.a. | 6.200 |
| | SD | (0.772) | (0.860) | n.a. | (0.788) | (2.293) | (2.340) | n.a. | (2.151) |
| | N | 56 | 57 | 0 | 40 | 56 | 57 | 0 | 40 |
| Jefferson | Mean | 0.089 | 0.348 | −0.036 | 0.245 | 5.205 | 5.412 | 5.236 | 5.887 |
| | SD | (0.848) | (0.927) | (0.763) | (0.945) | (2.398) | (2.592) | (2.359) | (2.729) |
| | N | 112 | 119 | 72 | 62 | 112 | 119 | 72 | 62 |
| Uxton | Mean | 0.250 | 0.666 | 0.254 | 0.718 | 5.747 | 6.493 | 6.061 | 7.213 |
| | SD | (0.751) | (0.826) | (0.775) | (0.792) | (2.174) | (2.212) | (2.269) | (2.245) |
| | N | 265 | 229 | 131 | 155 | 265 | 229 | 131 | 155 |
| Average | Mean | 0.195 | 0.534 | 0.150 | 0.540 | 5.583 | 6.072 | 5.765 | 6.735 |
| | SD | (0.729) | (0.831) | (0.802) | (0.827) | (2.218) | (2.393) | (2.324) | (2.422) |
| | N | 525 | 405 | 204 | 257 | 525 | 405 | 204 | 257 |

students in the treatment schools and is lower than the rest because this plot accounts for differences based both on English proficiency and treatment group status at the start of the study (solid thin line).

RQ1. How did English speaking students from English-language homes (ELH) who participated in the Word Generation program learn, maintain, and consolidate

words compared with similar students attending comparison schools?

Each of the parameter estimates for student learning, maintenance, and consolidation that do not invoke LMH status or LEP status specify the average scores to English-proficient students from English-speaking homes. In the final fitted model both treatment and comparison

Table 5. Average vocabulary scores on all eleven longitudinal items by wave by language status.

| | | | Scaled score | | | | Raw score | | | |
|------------|-----|---------|--------------------|-------------|----------------|-------------|--------------------|-------------|----------------|-------------|
| | | | Instructional Year | | Follow-up year | | Instructional year | | Follow-up year | |
| | | | Fall 2007 | Spring 2008 | Fall 2008 | Spring 2008 | Fall 2007 | Spring 2008 | Fall 2008 | Spring 2008 |
| Comparison | LMH | Non LEP | 0.310 | 0.679 | 0.317 | 0.699 | 5.915 | 6.526 | 6.204 | 7.193 |
| | | | (0.717) | (0.799) | (0.759) | (0.733) | (2.115) | (2.197) | (2.197) | (2.197) |
| | | LEP | -0.404 | 0.056 | -0.224 | 0.128 | 3.909 | 4.800 | 4.704 | 5.545 |
| | | | (0.600) | (0.621) | (0.563) | (0.681) | (1.646) | (1.750) | (1.728) | (2.087) |
| | EH | Non LEP | 0.203 | 0.521 | 0.159 | 0.524 | 5.601 | 5.996 | 5.810 | 6.698 |
| | | | (0.778) | (0.922) | (0.811) | (0.924) | (2.250) | (2.523) | (2.416) | (2.580) |
| | | LEP | -0.139 | 0.069 | 0.164 | -0.019 | 4.600 | 4.800 | 5.500 | 4.667 |
| | | | (0.733) | (0.559) | (0.561) | (0.621) | (2.074) | (1.095) | (0.707) | (1.528) |
| Treatment | LMH | Non LEP | 0.058 | 0.639 | 0.236 | 0.676 | 5.070 | 6.395 | 5.968 | 7.120 |
| | | | (0.701) | (0.760) | (0.750) | (0.716) | (2.055) | (2.169) | (2.243) | (2.118) |
| | | LEP | -0.466 | -0.040 | -0.453 | 0.030 | 3.524 | 4.476 | 3.855 | 5.273 |
| | | | (0.534) | (0.845) | (0.760) | (0.787) | (1.581) | (2.273) | (2.039) | (2.281) |
| | EH | Non LEP | -0.113 | 0.381 | 0.038 | 0.389 | 4.638 | 5.765 | 5.472 | 6.445 |
| | | | (0.744) | (0.832) | (0.797) | (0.847) | (2.175) | (2.327) | (2.384) | (2.482) |
| | | LEP | -0.565 | -0.203 | -0.612 | 0.074 | 3.600 | 4.385 | 3.500 | 5.600 |
| | | | (0.423) | (0.605) | (0.597) | (0.776) | (1.506) | (1.805) | (1.716) | (2.302) |

students from English-speaking homes made wave-to-wave improvement in their vocabulary knowledge ($\gamma_{10} = 0.371, p < .001$). To be sure that this estimate was not unduly influenced by the large school-level differences, we fit the final model with a set of dummy variables and found that the effect of instruction was stable. Both treatment and comparison students also experienced a summer setback, which is defined as the difference between their vocabulary score after summer vacation and the score we would have expected if they had continued to learn at a constant rate through the year ($\gamma_{30} = -0.639, p < .001$). Treatment students from English-speaking homes also experienced a one-time improvement at the end of the instructional period ($\gamma_{20} = 0.169, p < .001$), which they maintained compared with comparison students during the study.

These results are clearly visible in Figure 1. The bold dashed line (second from the top) represents the trajectory of typical sixth-grade students from English-speaking homes who are not in the Word Generation program. The heavy solid line (fourth from the top) presents the trajectory of prototypical sixth-grade students from English-language homes in the treatment schools. These students have steeper trajectories during the year of instruction, significantly narrowing the gap between themselves and comparison students. Interestingly, after the instructional period the trajectories of treatment and comparison students are completely parallel, suggesting

no relative loss of word knowledge by treatment students even a year after instruction.

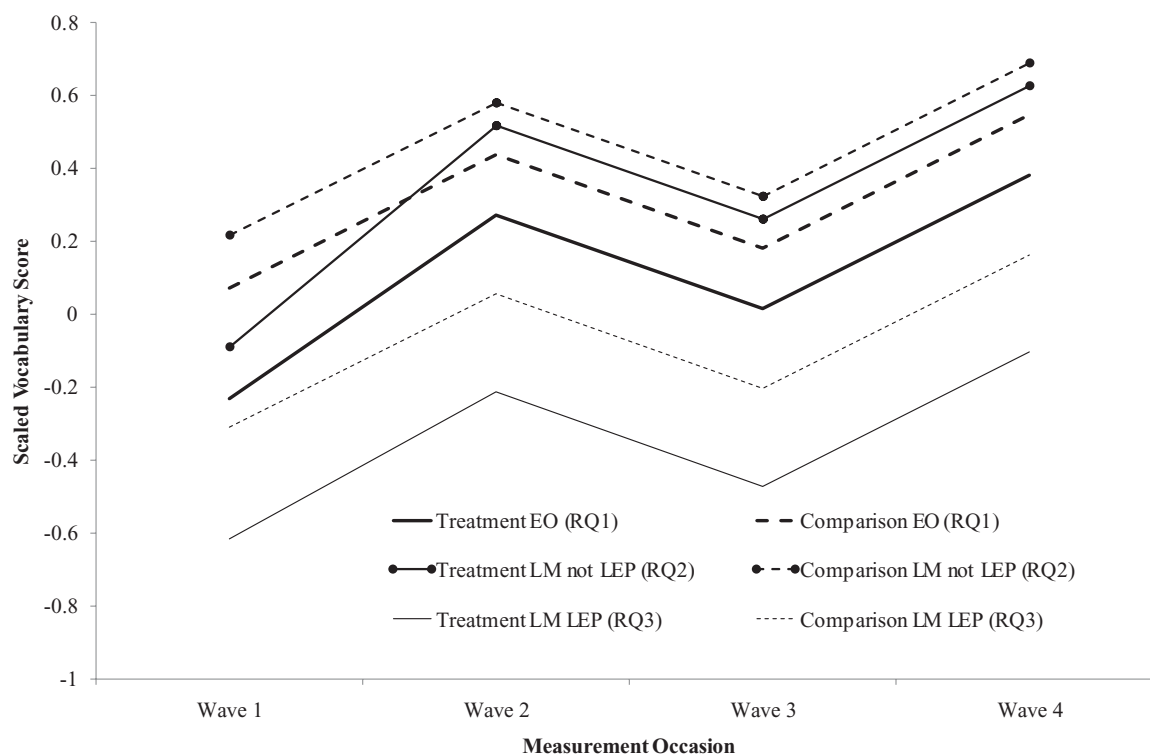
RQ2. How did English-proficient students from language-minority homes (LMH) who participated in the Word Generation program learn, maintain, and consolidate words compared with similar students attending comparison schools?

At the start of the study English-proficient students from language-minority homes had better scaled vocabulary scores on average than English-proficient students from English-only homes ($\gamma_{04}LMH_i = 0.138, p < .001$), although they experienced the same growth and summer setback as students from English homes ($\gamma_{14}LMH_i$ and $\gamma_{32}LMH_i$ were not significant and are not reported in the final fitted model). English-proficient students from language-minority homes who participated in the Word Generation program benefited even more than students from English homes ($\gamma_{23}LMH_i = 0.107, p < .01$). These results can be seen clearly in Figure 1. English-proficient students from language-minority homes from the comparison group (dashed line with marker) started the study with stronger vocabulary scores than English-proficient students from language-minority homes attending treatment schools (solid line with markers). However, during the instructional period LMH students in the treatment schools made strong gains, ending the

Table 6. *Taxonomy of multilevel models for change predicting VOCAB across four waves of data.*

| | | Relevant research question | Parameter | Model A | Model B | Model C |
|-------------------------------|-----------------------|----------------------------|-----------------|----------------------|----------------------|----------------------|
| Level 1 predictors | Intercept | | γ_{00} | 0.246*** (0.016) | 0.081*** (0.018) | 0.103** (0.035) |
| | WAVE | | γ_{10} | | | 0.371*** (0.018) |
| | SUMMER | | γ_{30} | | | -0.639*** (0.035) |
| | GR8*INSTR | | γ_{22} | | | -0.136** (0.045) |
| Level 2 predictors | GR8 | | γ_{02} | | 0.154*** (0.007) | 0.297*** (0.039) |
| | WG_SCHOOL | | γ_{03} | | | -0.309*** (0.038) |
| | LMH | | γ_{04} | | | 0.138*** (0.039) |
| | LEP | | γ_{05} | | | -0.526*** (0.067) |
| Question predictors | INSTRUCTION | RQ1 | γ_{20} | | | 0.169*** (0.031) |
| | LMH by INSTRUCTION | RQ2 | γ_{23} | | | 0.107** -0.039 |
| | LEP by INSTRUCTION | RQ3 | γ_{24} | | | -0.205** (0.068) |
| Level 1 Variance Component | Residual | | ϵ_{ij} | 0.3189*** (0.008) | 0.2756*** (0.007) | 0.238*** (0.006) |
| Level 2 Variance Component | Intercept | | ξ_{0i} | 0.3786*** (0.017) | 0.4161*** (0.018) | 0.372*** (0.015) |
| Goodness-of-fit | -2 LL | | | 11403.3 | 11009.5 | 10299.0 |

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



EO = English only; LEP = limited English-proficiency; LM = language minority; RQ1 = research question 1; RQ2 = research question 2; RQ3 = research question 3
 Figure 1. Prototypical plot of sixth-grade students in treatment and comparison groups by language status.

study with significantly improved scores. A post hoc GLH tests demonstrated that there was no difference between English-proficient students from language-minority homes in the treatment and comparison schools at the end of the instructional period ($\chi^2 = 0.52, ns$).

RQ3. How did students with limited English proficiency (LEP) from language-minority homes who participated in the Word Generation program learn, maintain, and consolidate words compared with similar students attending comparison schools?

LEP students in both the treatment and comparison schools started the study with lower vocabulary skills ($\gamma_{05}LEP_i = -.526, p < .001$), and experienced the same growth and summer setback as students from English-speaking homes (the terms $\gamma_{15}LEP_i$ and $\gamma_{33}LEP_i$, were not significant and are not reported in the final fitted model). An interaction between language proficiency and instruction ($\gamma_{24}LEP_i = -0.205, p < .001$) was negative, eliminating the predicted benefit of instruction ($\gamma_{20} = 0.169, p < .001$). Since there was no overall predicted improvement for LEP students participating in Word Generation, we should not see any difference between gains by students in treatment and comparison schools. GLH tests proved there were no differences in the growth of these groups during the instructional period ($\chi^2 = 0.23, ns$). These results are evident in Figure 1: the vocabulary-

learning trajectories of treatment and comparison LEP students are parallel across the course of the study.

Discussion

In most respects the findings from this study are congruent with the previous evaluations of the Word Generation program. During the intervention period, treatment students made significant gains relative to students in the comparison school on average. Furthermore, gains were larger for English-proficient LMH students than for students from English-speaking homes (Snow et al., 2009). The current study allowed us to examine the long-term effect of program participation on student vocabulary for ELH, LMH and LEP students. English-proficient students from language-minority homes who participated in the program made strong gains and maintained them compared to comparison students even a year later. English proficient students from English-speaking homes also made gains relative to the comparison group and maintained those gains across the course of the study. LEP students, however, did not show short-term or long-term benefits from participation in the Word Generation program.

These data reinforce the findings of Kieffer (2008) and Uchikoshi (2006): there are large differences between proficient students from language-minority homes and students who enter school with limited

English proficiency. Our findings supplement and extend Kieffer's (2008) analysis of K-5 students using a nationally representative dataset. Kieffer found a small advantage for language-minority students who were English proficient when entering school and a large deficit for students who entered school with limited English proficiency. Our findings suggest that students from language-minority homes, whether formerly limited in English proficiency or not, still show vocabulary deficits, but that such deficits can be addressed instructionally. Those still classified as LEP in the middle grades, however, continue to lag in vocabulary even after receiving targeted instruction.

LEP treatment students in this study did no worse or better than students in the comparison schools, suggesting a disparity between the program and the needs and capacities of these learners. We have several ideas about which aspects of the program could be improved for such students and are working on adaptations. First, although the target words were selected as ones that students would regularly encounter in text and in their content-area instruction, it is possible that LEP students had insufficient exposure to these words outside their 15 minutes of Word Generation instruction. Given that adolescent vocabulary development can be supported by independent reading (Fukkink, Blok & de Gloppe, 2001; Lawrence, 2009), LEP students may have been disadvantaged because they were not assigned or could not access grade-level texts that used academic words. Second, considering how low the scores of sixth-grade LEP students were, it is probable that the target words were too difficult for these students. Indeed, academic English is cognitively demanding for all students (Scarcella, 2003). However, while English proficient students could direct their capacities toward conceptual and vocabulary development, LEP students were simultaneously learning the phonological, grammatical, and pragmatic features of English in the process.

The high cognitive load is compounded by features of the curriculum. LEP students who received no L1 language support to foster language development may have found that the materials were too difficult. We are currently creating a new curriculum devoted to supporting ELLs based on research-based recommendations for instruction and academic interventions (Francis, Rivera, Lesaux, Kieffer & Rivera, 2006). This curriculum incorporates elements that have been shown to be effective with other samples of language-minority learners, such as building on cognate knowledge (e.g., August et al., 2009; Carlo et al., 2004; Townsend & Collins, 2009).

On the other end of the spectrum, eighth-grade students from English-speaking homes did not benefit from the program, and while a post hoc GLH test shows that proficient eighth-graders from LM homes did benefit from program participation ($\chi^2 = 7.53$, $p = .006$), the improvement of these older students was reduced.

These data suggest that while the words chosen for this curriculum may have been too hard for some students, they may have perhaps been too easy for others. This does not necessarily mean that the curriculum was not challenging. Much of the actual Word Generation program is focused on providing opportunities for discussion and writing persuasively about a topic, tasks which require many academic language skills to complete. However, these data do suggest more challenging words would create greater learning opportunities for older students.

In addition to increasing our understanding of how children learn academic vocabulary in a second language, these results also provide us with guidance about how we can improve the Word Generation program and our work with schools and school districts. This work was driven by a district identified problem and the curriculum was created in close collaboration with teachers. While there is ample research to suggest that academic vocabulary is tightly connected to reading ability, especially in later elementary and middle grades, we think it is critical that vocabulary was a topic that our collaborating teachers identified as a high priority in interviews and surveys; we consider it essential that we as a research community find ways to include teachers' perspectives in deciding what education research should be conducted if we expect research to influence practice. Our approach to ongoing analysis of student outcomes using longitudinal data allowed us to interpret our results within the messy context of student learning during the summer and school year, and to maximize our data by comparing gains associated with program participation to gains in both treatment schools (in the follow-up year) and comparison schools. We are optimistic that longitudinal research methods that examine the value-added effect of program participation (Biancarosa, Bryk & Dexter, 2010) will allow more collaborative relationships between school districts and researchers working to develop and evaluate instructional interventions and approaches.

Limitations and future research

There are several limitations to the study. During the first year pre-tests were not administered at treatment and control schools at the same time (as discussed in Snow et al., 2009). Additionally, as mentioned, the treatment and control schools were not well matched, nor do we have good measures of fidelity of implementation. We did not have sufficient power to examine difference in vocabulary maintenance at the school level. Due to the changing of teachers across grades, we did not model the cross-classification of students by teachers. It is possible that classroom-level variability due to instruction and grouping of students may have interesting implications for examining program implementation and effects. Our

plans for future research include testing the effects of Word Generation as a randomized field trial and more closely monitoring implementation.

Although the current study contributes to the literature by examining the impact of instruction for students of various home-language statuses and proficiency levels, the policy driven language proficiency descriptors are nonetheless broad and rough. Thus, future research should continue to examine the impact of intervention on students by home language, but use proficiency scales based on English achievement measures instead of these rough categories.

Our vocabulary measure is a multiple-choice task that requires participants to choose synonyms. Although this measure is easy to administer in a whole group setting, it is not as complete a measure of vocabulary depth and knowledge as we would like; knowledge of the distractors can be confounded with knowledge of target words. Additionally, it allows us to determine how well students maintained or consolidated their receptive vocabulary, but provides no indication of their productive word knowledge. Our ongoing studies of the Word Generation program use several assessments of depth of vocabulary knowledge. Although these assessments will help us better understand how various kinds of semantic knowledge relate to learning and maintenance, preliminary results show that while our generic multiple choice tests are not sophisticated, they are reliable and highly correlated with a range of other measures of target word knowledge.

Despite these limitations, the current study makes two noteworthy contributions to the research base of vocabulary interventions with language-minority and English-proficient students. First, it highlights the importance of examining the impact of instruction for students of various home language statuses and language proficiencies. Only by distinguishing proficient and limited proficient students from language-minority homes were we able to understand the unique needs of the latter group and make program adjustments. Second, it indicates vocabulary instruction can result in robust learning for proficient students from language-minority homes, learning that is as stable as the vocabulary knowledge garnered through multiple incidental exposures in text and discussion typical in non-intervention school settings. We take these findings as support for an approach to vocabulary instruction that emphasizes the contextualized use of words in multiple academic contexts and in multiple modalities, and emphasize the use of high leverage academic language in discussion and debate. While this approach did not result in improvement for all students, those students that benefited from these activities, especially English proficient students from language-minority homes, demonstrated the effects of participation of the Word Generation program even a year after instruction.

References

- Alexander, K., Entwisle, D., & Olson, L. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation & Policy Analysis*, 23 (2), 171–191.
- Alexander, K., Entwisle, D., & Olson, L. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, 72(2), 167–180.
- Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K., Fox, M., Frohlich, L., Kemp, J., & Drake, L. (2010). *The condition of education 2010* (NCES 2010–028). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness*, 2 (4), 345–376.
- August, D., & Shanahan, T. (2006). Synthesis: Instruction and professional development. In D. August & T. Shanahan (eds.), *Developing literacy in a second language: Report of the National Literacy Panel*, pp. 351–364. Mahwah, NJ: Lawrence Erlbaum.
- Bardovi-Harlig, K., & Stringer, D. (2010). Variables in second language attrition. *Studies in Second Language Acquisition*, 32 (1), 1–45.
- Beck, I., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.
- Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *Elementary School Journal*, 111 (1), 7–34.
- Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., Livey, T. J., & White, C. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, 39 (2), 188–215.
- Carver, R. (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., & Jacobs, V. (2003). Poor children's fourth-grade slump. *American Educator*, 27 (1), 14–17.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213–238.
- de la Fuente, M. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research*, 10 (3), 263–295.
- Entwisle, D., Alexander, K., & Olson, L. (1997). *Children, schools and inequality*. Boulder, CO: Westview Press.
- Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- Fukkink, R., Blok, H., & de Glopper, K. (2001). Deriving word meaning from written context: A multicomponential skill. *Language Learning*, 51 (3), 477–496.

- Fukkink, R., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from contexts: A meta-analysis. *Review of Educational Research*, 68 (4), 450–469.
- Heyns, B. (1978). *Summer learning and the effects of schooling*. New York: Academic Press.
- Kieffer, M. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology*, 100 (4), 851–868.
- Lawrence, J. F. (2009). Summer reading: Predicting adolescent word learning from aptitude, time spent reading, and text type. *Reading Psychology*, 30 (5), 445–465.
- Lawrence, J. F. (in press). English vocabulary learning trajectories of students whose parents speak a language other than English: Steep learning and deep summer setback. *Reading and Writing: An Interdisciplinary Journal*, doi: 10.1007/s11145-011-9305-z. Published online by Elsevier, March 27, 2011.
- Lawrence, J. F., White, C., & Snow, C. E. (2010). The words students need. *Educational Leadership*, 68 (2), 22–26.
- Lutkus, A. D., Rampey, B. D., & Donahue, P. (2005). *The nation's report card: Trial urban district assessment reading 2005* (NCES 2006-455). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Min, H. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning*, 58 (1), 73–115.
- Muthén, L. K., & Muthén, B. O. (2007). *Mplus: Statistical analysis with latent variables*. Los Angeles, CA: Muthén & Muthén.
- Nagy, W., & Scott, J. A. (2000). Vocabulary processes. In M. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (eds.), *Handbook of reading research* (vol. III), pp. 269–284. Mahwah, NJ: Lawrence Erlbaum.
- National Institute of Child Health and Human Development [NICHD]. (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* (NICHD 00-4769). Washington, DC: U.S. Government Printing Office.
- Proctor, C., Dalton, B., Uccelli, P., Biancarosa, G., Mo, E., Snow, C. E., & Neugebauer, S. (2009/2011). Improving comprehension online: Effects of deep vocabulary instruction with bilingual and monolingual fifth graders. *Reading and Writing: An Interdisciplinary Journal*, 24 (5), 517–544. [Online publication 2009, print publication 2011.]
- Scarcella, R. (2003). Academic English: A conceptual framework. http://www.lmri.ucsb.edu/publications/03_scarcella.pdf (retrieved October 20, 2008).
- Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and even occurrence*. New York: Oxford University Press.
- Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, 2 (4), 325–344.
- Suárez-Orozco, C., Suárez-Orozco, M., & Todorova, I. (2008). *Learning a new land: Immigrant students in American society*. Cambridge, MA: Belknap Press.
- Townsend, D., & Collins, P. (2009). Academic vocabulary and middle school English learners: An intervention study. *Reading and Writing: An Interdisciplinary Journal*, 22 (9), 993–1019.
- U.S. Department of Education. (2001). No child left behind. <http://www.ed.gov/nclb/landing.jhtml> (retrieved May, 2004).
- Uchikoshi, Y. (2006). English vocabulary development in bilingual kindergartners: What are the best predictors? *Bilingualism: Language and Cognition*, 9 (1), 33–49.
- Vaughn, S., Martinez, L., Linan-Thompson, S., Reutebuch, C., Carlson, C., & Francis, D. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness*, 2 (4), 297–324.
- Verhallen, M., & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Applied Linguistics*, 14 (4), 344–363.