Read Writ (2017) 30:829–856
DOI 10.1007/s11145-016-9703-3

# Defying expectations: Vocabulary growth trajectories of high performing language minority students

**Jin Kyoung Hwang**[1] · **Joshua Fahey Lawrence**[1,2] · **Catherine E. Snow**[3]

Published online: 22 October 2016


**Abstract** We investigated general vocabulary and academic vocabulary growth trajectories of adolescent language minority students using an individual growth modeling approach. Our analytical sample included 3161 sixth- to eighth-grade students from an urban school district in California. The language minority students in our sample were classified as initially fluent English proficient (IFEP), redesignated fluent English proficient (RFEP), or limited English proficient (LEP) students. The analytical sample was not a nationally representative sample and included a great number of Asian students and students who receive gifted and talented education. Students were assessed at four time points on a standardized measure of general vocabulary and a researcher-developed academic vocabulary test. On both vocabulary measures, IFEP students slightly outperformed English-only (EO) students on average, and EO students scored higher than RFEP and LEP students at baseline. RFEP and LEP students showed slower rate of growth than their EO peers in general vocabulary. While both EO and language minority students showed summer setback with general vocabulary knowledge on average, the magnitude of summer setback was not as great for LEP students. In academic vocabulary, all subgroups of language minority students showed more rapid rate of growth than their EO peers. Only the REP students experienced a change in the learning trajectory during the summer months. We discuss the implications of these findings for all language groups.

✉ Jin Kyoung Hwang
jkhwang1@uci.edu

1 School of Education, University of California, 3200 Education Building, Irvine, CA 92697-5500, USA

2 Department of Educational Research, University of Oslo, P.O Box 1092, Blindem, 0317 Oslo, Norway

3 Harvard Graduate School of Education, Larsen 313, 13 Appian Way, Cambridge, MA 02138, USA

**Keywords** Language minority students · Adolescent · Vocabulary growth · Longitudinal · Academic language

## Introduction

There are a large and growing number of language minority (LM) students in American public schools. In 2011, over 60 million Americans over the age of five spoke a language other than English at home (Ryan, 2013), and that number is projected to increase in coming years (Shin & Ortman, 2011). LM students, like their English-speaking peers, may struggle with reading. In early grades these students need to know how to phonologically represent a new language, understand a new system for representing sounds, and comprehend the decoded text. Fortunately, the vocabulary demands in early grade texts are low: most words in primary grade texts are high-frequency words that are semantically transparent. Texts in middle school are more complex. In secondary schools, children are expected to master abstract vocabulary words in reading (Leach, Scarborough, & Rescorla, 2003; Short & Fitzsimmons, 2007; Storch & Whitehurst, 2002). The Common Core State Standards (CCSS) may intensify the challenge LM students face with the increased emphasis on nonfiction texts, academic discussion, and argumentation (Bunch, Walqui, & Pearson, 2014; Common Core State Standards Initiative, 2010). We must understand the developmental trends in LM students' general language and academic language proficiency if we are to support them. In this study we model students' general and academic vocabulary growth across middle grades, a time when the proportion of academic words appearing in texts increases and knowledge of them is of critical importance for academic success (Corson, 1997; Nagy & Townsend, 2012; Townsend, Filippini, Collins, & Biancarosa, 2012).

### Language designations

In this paper, we use the term *language minority student* to describe any school-aged student whose home language is not English (August & Shanahan, 2006). Across the U.S., more than 300 languages are being used in LM students' homes (National Clearinghouse for English Language Acquisition, 2011). Spanish is the most-spoken home language in many states in the U.S. followed by Vietnamese, Chinese (including all Chinese dialects), Arabic, and Hmong. Great numbers of school-aged LM students are U.S.-born citizens (82 %) and approximately 30 % of this population are from families whose income levels are below the poverty threshold (Aud et al., 2011). Students from LM homes may vary on many other dimensions such as their proficiency in their home language, proficiency in English, schooling history, and so forth. Within this extremely large and heterogeneous group, policy makers have identified three relevant subgroups based on their English proficiency (Ragan & Lesaux, 2006): initially fluent English proficient (IFEP), limited English proficient (LEP), and redesignated fluent English proficient (RFEP).

IFEP students are those who enter school with enough English proficiency and who do not receive additional language support or assessment. LEP students [also commonly known as English language learners (ELLs)] do not have the sufficient English proficiency to function in schools and receive additional support from the schools they attend. RFEP students are those who were identified as LEP when they first entered school, but subsequently met proficiency criteria and were redesignated by their school (at which point they stopped receiving support).

Because there is no specific guideline at the federal level for identifying and classifying LM students, state- and district-level proficiency thresholds largely determine whether a student is identified as IFEP, RFEP, or LEP (Bailey & Kelly, 2013; Linquanti & Cook, 2013; Ragan & Lesaux, 2006). One empirical analysis of data from three states shows that the more stringent the classification criteria, the larger the difference between RFEP and LEP students is (Kim & Herman, 2009). Kim and Herman also found that in the state with the highest standards for redesignation, RFEP students actually outperform English-only (EO) students on some measures. However, one consequence of a stringent policy may be that some students are never redesignated as being English proficient.

The data for the study presented here come from a large urban district in California. The state-level guidelines in California prescribe that LEP students be redesignated based on their scores on an English-proficiency assessment, academic content-area assessment, teacher evaluation, and parental opinion (English Language Proficiency Assessment of 1999, 2014). In this particular school district, LEP students needed to obtain either *early advanced* or *advanced* on the California English Language Development Test (CELDT), which assesses students' overall English proficiency, as well as *proficient* or *advanced* on the English language arts subtest of the California Standards Test (CST, www.cde.ca.gov). Up until the 2013–2014 school year, all students in California took the CST for the first time at the end of second grade. Therefore, LEP students could be reclassified as early as the middle of third grade when the second grade CST scores were available. With the implementation of Smarter Balanced Assessments (www.smarterbalanced.org) during the 2013–2014 academic year, the state of California no longer administers CSTs to their students. Thus, the redesigntion process does not use CST results anymore. LEP students were required to take the CELDT annually until they are redesignated. Both federal and California state law require that LEP students receive some kind of services/resources for their English language development. Although these state-level guidelines set a high bar for LEP students to be reclassified (i.e., satisfying all the requirements at once), the ultimate decision is made by personnel in each district. California's stringent reclassification criteria may be one of the reasons why the majority (approximately 60 %) of California secondary school LEP students are long-term LEPs (Olsen, 2010; Parrish, Perez, Merickel, & Linquanti, 2006).

## LM students' vocabulary knowledge

There is a strong relationship between second language (L2) vocabulary knowledge and reading comprehension (August et al., 2005, Proctor et al., 2005; Hoover &

Gough, 1990; Lesaux, Lipka, & Siegel, 2006; Rydland, Aukrust, & Fulland, 2012). Although vocabulary knowledge is one of the foundational skills for all students who *read to learn* new concepts across different content areas (Chall & Jacobs, 2003), national statistics and research studies show that LM students (particularly those with limited English proficiency) tend to lag behind their EO counterparts in English vocabulary knowledge (August & Shanahan, 2006; National Center for Education Statistics, 2012). Cummins' (1979) threshold hypothesis suggests we will find differential vocabulary growth trajectories for LM students with different English proficiency levels. Cummins (1979) suggests that there are "thresholds" that students must meet in order to avoid any disadvantage and enjoy benefits that are associated with bilingualism. In other words, students with low levels of their first and second language proficiency (i.e., semilingualism) are likely to experience academic disadvantages from insufficient interaction in their educational environments. Furthermore, students who are proficient in either language (i.e., dominant bilingualism) are not likely to experience such disadvantage, but students who are fully proficient in both languages (i.e., additive bilingualism) are likely to benefit from cognitive and academic advantages that are associated with bilingualism. However, with threshold hypothesis, it is not clear whether attaining a certain threshold affects LM students' general word learning ability or differentially affects word-learning skills in different contexts.

Academic words are usually classified as general or discipline-specific academic words (Hiebert & Lubliner, 2008; Nagy & Townsend, 2012). The type of academic vocabulary that is explored in this study is general academic vocabulary. *General academic words* have been hypothesized to be particularly important for academic reading (Snow & Uccelli, 2009; Nagy & Townsend, 2012). They are the glue words of academic texts. They appear frequently in texts across content areas but are not tied with any discipline-specific content or idea. Coxhead (2000) created the Academic Word List using both word frequency thresholds (i.e., the most frequent words in English were excluded) and dispersion thresholds (i.e., only words which appeared regularly across academic disciplines were included). Researchers have argued that because these words appear across academic disciplines with some regularity, knowledge of them is particularly useful for LEPs (Baumann & Graves, 2010; Coxhead, 2000; Nagy & Townsend, 2012). We hypothesized that students could learn and be exposed to general vocabulary words not only in school settings but also in other non-academic settings (e.g., home environment, media) whereas the context for learning academic vocabulary words can be more limited to school settings. We believed that examining general and academic vocabulary separately would provide insight about how LM students develop their vocabulary knowledge across proficiency thresholds.

### *General vocabulary knowledge*

There are few longitudinal studies that have investigated preadolescent and adolescent EO and LM students' general vocabulary growth (Lawrence, 2012; Jean & Geva, 2009; Mancilla-Martinez & Lesaux, 2011a, b). These studies show that LM students may have different vocabulary growth trajectories than their EO peers during their elementary and middle school years. Lawrence (2012) examined the English vocabulary learning trajectories of sixth- and seventh-grade EO ($n = 210$) and LM ($n = 68$) students from an urban school district in the Northeast that served large numbers of minority students and students from low socioeconomic status (SES) homes. LEP students were not included in his LM student sample. By collecting data across two school years including one summer break, Lawrence (2012) was able to examine students' vocabulary growth not only during the academic year but also during summer when the usual school instruction does not happen. This is of particular importance as previous research has shown students, especially those from low-SES environment, experience summer setback with their academic outcomes (Alexander, Entwisle, & Olson 2001; Heyns, 1978). Using four waves of scores from the Group Reading Assessment and Diagnostic Evaluation (Williams, 2001), he found that all students improved during the school year but experienced roughly a two-month summer setback in their vocabulary knowledge. Moreover, LM students regressed in vocabulary knowledge more severely than their EO peers during summer, but had steeper vocabulary learning trajectories during the school year. This pattern held even when well-known predictors of vocabulary (i.e., time spent on independent reading) and summer setback (i.e., SES) were controlled for.

Mancilla-Martinez and Lesaux (2011b) modeled growth rates of Spanish-speaking LM students' ($N = 173$) word reading and vocabulary knowledge in English and Spanish from ages 4.5 to 11. In this study, the researchers did not classify LM learners into different subgroups. The participants were first recruited when they were enrolled in either Head Start programs or public preschool programs in the Northeast. Researchers assessed students' word reading and vocabulary knowledge with the Letter-Word Identification subtest and Expressive Vocabulary subtest from the Woodcock Language Proficiency Battery-Revised (Woodcock, 1991), respectively. They found that LM students' word reading growth rate and ability were on par with the national monolingual sample used to norm the test, but their vocabulary knowledge was consistently below the national norm.

### *Academic vocabulary knowledge*

Some components of the L2 vocabulary develop rapidly, while others, such as knowledge of abstract or academic terms, grow slowly (August & Shanahan, 2006; Collier, 1989; Cummins, 2008; Goldenberg, 2010; Mancilla-Martinez & Lesaux, 2011a). In their study, Lawrence, Capotosto, Branum-Martin, White, and Snow (2012) collected four waves of academic vocabulary data from treatment and comparison students participating in a quasi-experimental study of an academic language intervention, Word Generation. This study was conducted in Boston Public Schools that served LM students (38 %) and students from low-income homes (74 %). Approximately 20 % of the student population in Boston was classified as being LEP at the time this study was conducted. Using 11 anchor items embedded within each wave of testing, they created an item response theory (IRT) model that formed a time-varying level-1 outcome. Baseline scores indicated that LEP students had lower academic vocabulary scores than both EO students and English-proficient LM students. Interestingly, English-proficient LM students outperformed EO students at baseline. Although there were differences in response to program participation by language status, the growth trajectories of all groups of students in the comparison schools were the same; in other words, there was no difference in the rates that different groups of middle school students learned academic words in "business-as-usual" schools.

In a recent cross-sectional study, Uccelli, Galloway, Barr, Meneses, and Dobbs (2015) found that there was only a small difference between English-proficient students (including both EOs and IFEPs) and RFEP students on the researcher-developed Vocabulary Association Test (see Lesaux, Kieffer, Faller, and Kelley, 2010 for more information on this assessment) across grade levels (+.2 SD favoring English-proficient students in fourth grade, +.39 SD favoring RFEP students in fifth grade and no difference in sixth grade). The sample characteristics were not too different from other studies that were conducted on LM students. The majority of the sample in this study also came from socioeconomically disadvantaged homes (65 %). In their study, Lesaux et al. (2010) found that although LM control students had lower academic vocabulary pretest scores than EO control students on average, they improved roughly one point by the end of the study, whereas the EO control students did not improve. Because of state-by-state redesignation policies, it is not possible to draw definite conclusions about baseline differences in academic vocabulary for language groups across studies. However, these studies suggest that proficient LM students may learn academic words as quickly as, or more quickly than, EO students even though their baseline scores might be lower. These suggestive findings warrant further investigation.

LEP students' vocabulary knowledge is lower than their non-LEP counterparts (National Center for Education Statistics, 2012), however, it is not clear how proficient adolescent LM students (i.e., IFEP and RFEP students) are faring on vocabulary outcomes in comparison to their EO peers. Nor is it clear if these differences are maintained over time, or are replicated in measures of student's academic vocabulary. As mentioned above, there are studies that suggest proficient LM students may show more growth compared to their EO peers during upper-elementary or middle school years (Lawrence et al., 2012; Uccelli et al., 2015). However, there is also evidence showing that a large proportion of reclassified students are still experiencing difficulties in English language arts and mathematics and a considerable number are retained one grade or more (Slama, 2014). It is worth exploring whether these academic challenges are associated with differences in students' academic vocabulary knowledge, given its centrality to skilled content area reading. Thus, the research questions that will be addressed in this study are the following:

1. What are the general vocabulary trajectories of middle school students in each language designation controlling for their student-level reading-related skills and characteristics?
2. What are the academic vocabulary trajectories of middle school students in each language designation controlling for their student-level reading-related skills and characteristics?

We fit a confirmatory factor model to these data to test if vocabulary was unidimensional construct across the general and academic vocabulary tests. Our results with these data indicate that it is.[1] Another analysis testing two factor models based on test type, and two factor models based on academic and non-academic words (regardless of which test they were in) using the Metropolis–Hastings Robbins–Monro algorithm to estimate the unknown parameters (Cai, 2010) indicated that the dimensionality of vocabulary knowledge construct as measure by the synonym task can best be described in terms of one factor (Lawrence, Kulesz, Hwang, & Francis, under review). However, given the instructional prominence of a certain class of general academic words sometimes called *tier 2 words* (Beck, McKeown, & Kucan, 2002), the prominence of these types of words in policy documents (National Governors Association, 2010), and evidence that knowledge of these words relates to content-area learning (Townsend et al., 2012), there are theoretical and practical reasons to explore academic word learning rates as a function of student language status and word usage. We thought students could learn and be exposed to general vocabulary words in both academic and non-academic settings, but the context for learning academic vocabulary words can be more limited to school settings. We probed this hypothesis by modeling growth on two distinct vocabulary measures. While our general vocabulary measure included a wide range of English words, the words that were tested in academic vocabulary assessment tended to be mid-frequency words that are often encountered in texts across content areas. More information about how we operationalized general and academic vocabulary constructs are explained in the Measures section.

## Methods

### Sample

The data for this study came from control school students in an IES-funded randomized efficacy trial of Word Generation (www.wordgeneration.org) that took place from 2010 to 2012. Six- to eighth-grade students from six middle schools in a large urban district in California participated the study. Initially, there were 3653 students in this longitudinal study who contributed to at least one wave of data collection. Of these students, there were 474 who did not complete the baseline assessments and 96 students who were missing language proficiency status data that were essential in our analysis; in the end, our analytical sample included 3161 students. In Year 2, 1099 eighth-grade students left the study (i.e., graduated middle school).

Table 1 presents demographic information of the participating students by language designation (first four rows) and for the whole analytic sample (last row). Across the sample, 84 % of students were eligible for free or reduced lunch, but there were differences across groups. Only 71 % of EO students were eligible, while 92 % of LEP students were eligible. A very small number of RFEP (2 %) and IFEP

[1] Our confirmatory factor analysis with items from two measures indicated that the one-factor model (AIC = 661804, BIC = 663751, RMSEA = 0.030) fits the data better than the two-factor model (AIC = 664192, BIC = 666112, RMSEA = 0.032).

**Table 1** Demographic information of the participants

| | Eligible for FRL (%) | Special education (%) | Gifted and talented education (%) | Race | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Asian (%) | Hispanic (%) | Black (%) | White (%) | Other (%) |
| EO ($n = 948$) | 71 | 17 | 35 | 33 | 11 | 26 | 16 | 14 |
| IFEP ($n = 322$) | 80 | 4 | 51 | 61 | 22 | 2 | 5 | 10 |
| RFEP ($n = 1235$) | 90 | 2 | 58 | 77 | 17 | 0 | 2 | 4 |
| LEP ($n = 659$) | 92 | 20 | 9 | 52 | 39 | 1 | 2 | 7 |
| Total ($N = 3161$) | 84 | 10 | 40 | 57 | 20 | 8 | 6 | 8 |

*FRL* free or reduced lunch, *EO* English-only, *IFEP* initially fluent English proficient, *RFEP* redesignated fluent English proficient, *LEP* limited English proficient

(4 %) students had individualized education plans, and there was a large number of students in the gifted and talented education program (40 % across all students). The analytical sample consists of Asian (57 %), Hispanic (20 %), White (6 %), and Black (8 %) students. Most EO students were Asian (33 %) or Black (26 %). Students in the other language designations were mostly Asian and Hispanic. There were comparable percentages of students in each grade-level cohort: sixth-grade cohort (32 %), seventh-grade cohort (33 %), and eighth-grade cohort (35 %). Table 2 displays home language information of the participating LM students. The majority of the LM students in our sample were from Cantonese-Speaking homes. LM students from Spanish-speaking homes were the second largest group in our sample. Students' home language information was provided by the participating school district. All students are required to fill out the home language survey when they first enter the public school system in California. Although students' home language use can be dynamic and change as family members learn English, the district survey is administered only once and may be out of date.

As can be inferred from Tables 1 and 2, the sample in this study differs from the samples used in most published studies. In our sample, the majority of our LM students were Asians and there were a greater percentage of Hispanic students in the LEP group compared to other language groups. It is also worth noting that good number of IFEP and RFEP students were in the district gifted and talented program. Thus, while this sample was drawn from a large urban district in California [which serves more than 30 % of LEP students in the U.S. (National Center for Education Statistics, 2015)], their characteristics were distinct from those of the students attending other typical urban school districts in the U.S.

**Table 2** Language minority students' home languages

| | Cantonese | Spanish | Filipino | Vietnamese | Mandarin | Other |
|---|---|---|---|---|---|---|
| IFEP | 35 | 18 | 3 | 2 | 1 | 41 |
| RFEP | 47 | 15 | 3 | 3 | 2 | 30 |
| LEP | 25 | 32 | 7 | 3 | 2 | 31 |
| Total LM students | 40 | 20 | 4 | 3 | 2 | 31 |

*IFEP* initially fluent English proficient, *RFEP* redesignated fluent English proficient, *LEP* limited English proficient, *LM* language minority

## Measures

### *Time*

TIME is a level-1 variable indicating the time since the start of the study at which students completed the assessments. The data were collected in the fall and spring of two consecutive years. We coded each wave in months (i.e., wave 1 = 0 month, wave 2 = 7 months, wave 3 = 12 months, wave 4 = 19 months).

### *Language status*

The collaborating school district provided detailed information about participating LM students. Dummy variables for EO, IFEP, RFEP, and LEP students were used as student-level (level-2) predictors in our analyses. Although some of the LEP students' status changed during this course of the study (i.e., 229 LEP students in Year 1 were reclassified as RFEP in Year 2), we treated this variable as time-invariant, reflecting classification of students' English proficiency status at the beginning of the study (i.e., September Year 1). The analytical sample consists of 30 % EO, 10 % IFEP, 39 % RFEP, and 21 % LEP students.

### *General vocabulary*

The Gates-MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) is a group-administered assessment that includes vocabulary and passage-comprehension subtests. As the vocabulary subtest assesses "words of general usefulness," (MacGinitie et al., 2000, p. 9) the score from this test was used as an indicator of students' general vocabulary knowledge. The level 6 Form T was administered to sixth grade students and the level 7/9 Form T was given to seventh and eighth grade students across all data collection points. This test was administered twice each year, in September/October and in May. The vocabulary subtest consists of 45 multiple-choice items that ask students to choose the synonym of a target word. Extended scale scores of the vocabulary subtest (GV) were used in the analysis as they allow for estimating growth over time on a single scale (MacGinitie et al., 2000). These scores are scaled such that a score of 515 corresponds to average achievement at the beginning of sixth grade and 526 to seventh grade. The internal reliability (Cronbach's $\alpha$) for our analytical sample was .90 at the first wave. The mean and the standard deviation of the general vocabulary were 518.82 and 36.92, respectively, at the first wave.

### *Academic vocabulary*

Students' academic vocabulary was assessed with a 50-item multiple-choice test developed and validated by the Word Generation research team. For each item, the target word was embedded in a short sentence, and students were asked to choose a synonym for the target word from among four choices. The target words were taken from the Academic Word List (Coxhead, 2000). The same academic vocabulary test was administered twice, in September/October and in May. A different test form was used in the first (waves 1 and 2) and second (wave 3 and 4) year. Out of 50 items, 20 were anchor items (items that appear in both test forms), and 30 were unique to each test form. Academic vocabulary test items developed by the Word Generation research team can be found in the IRIS digital depository (www.iris-database.org).

To make responses from the two forms of the assessment equivalent, we conducted test scaling using IRT analysis. Scoring with IRT models allows students' academic vocabulary scores from each wave to be estimated on a common metric. We ran a unidimensional three-parameter IRT model on the first and third wave of data constraining the anchor items to have same item parameters (i.e., slope and intercept). In this process, we dropped seven anchor items, two unique items from the Year 1 test form, and one unique item from the Year 2 test form based on the overall fit index (i.e., root mean square error of approximation, RMSEA; Brown & Cudeck, 1993) and local dependence statistics. The final model fit the data well (RMSEA = .03). Marginal reliability for the first wave was .91 and that of the third wave was .92. Once item parameters were obtained, we used them to score all the waves of our data. The scoring method we used was *expected a posteriori* (EAP). These scores were used as an indicator for academic vocabulary (ACV) in our analysis. The scaled scores had a mean of −.02 and standard deviation of .89 at the first wave for our sample.

### *Reading comprehension*

Students' reading comprehension scores were used as a covariate in our analysis in order to control for baseline differences in reading among students. As vocabulary and reading comprehension skills are highly correlated constructs, it was important to control for students' initial reading-related skills in order to obtain more precise results. This covariate is not a time-varying variable and does not affect the estimation of students' growth over time.

The Gates-MacGinitie Reading Test (MacGinitie et al., 2000) was administered to assess students' reading comprehension skills. The level 6 Form T was administered to sixth grade students and the level 7/9 Form T was given to seventh and eighth grade students. Students were asked to read a passage and answer relevant comprehension questions (48 items total). Extended scale scores from the passage comprehension subtest (RC) obtained at the first wave were used as a covariate in our analysis to address both of our research questions. The Cronbach's $\alpha$ for our analytical sample was .91 in the first wave. The mean and standard deviation of our sample on reading comprehension was 524.40 and 37.79, respectively, at the first wave.

*Grade-level cohort*

To control for different grade levels in the analyses, student-level dummy variables were created for grades six (GRADE_6), seven (GRADE_7), and eight (GRADE_8). There was a comparable representation of sixth (32 %), seventh (35 %), and eighth (33 %) grade students in our sample.

*Summer*

How many summers (SUMMER) students experienced since the start of the study was also included in our analysis (e.g., wave 1 = 0, wave 2 = 0, wave 3 = 1, wave 4 = 1 for sixth- and seventh-grade cohorts). SUMMER was a time-varying continuous level-1 variable, and its parameter indicated whether or not students experience summer setback with their outcomes.

*Socioeconomic status (SES)*

Eligibility for receiving free or reduced lunch (FRL) was used as an indicator of students' SES. A student-level dummy variable was created to indicate students who received free or reduced lunch (FRL = 1) and those who did not (FRL = 0).

*Ethnicity*

Five student-level dummy variables (ASIAN, HISPANIC, BLACK, WHITE, and OTHER) were created to control for students' ethnicity. Asian students were used as the reference group in our analysis.

*Special education status*

A student-level dummy variable was created to indicate students who were receiving special education (SPED = 1) and those who were not (SPED = 0).

*Gifted and talented education status*

A student-level dummy variable was created to indicate students who were receiving gifted and talented education (GATE = 1) and those who were not (GATE = 0).

## Data analysis

To answer our research questions, we conducted multilevel models for change (Singer & Willett, 2003). These models allowed us to use all waves of data from each student to create a model of vocabulary growth over the course of two years. The data was prepared in a person-period dataset such that each student had up to four rows of data. The hypothesized multilevel model for change for the first research question was:

Level-1 (outcomes in four waves across 2 years):

$$\widehat{GV} = \pi_{0i} + \pi_{1i}TIME_{ij} + \pi_{2i}TIME_{ij}^{2} + \pi_{3i}SUMMER_{ij} + \varepsilon_{ij} \tag{1}$$

Level-2 (student-level):

$$\begin{aligned}\pi_{0i} = {} & \gamma_{00} + \gamma_{01}ACV_i + \gamma_{02}RC_i + \gamma_{03}IFEP + \gamma_{04}RFEP_i + \gamma_{05}LEP_i + \gamma_{06}GRADE7_i \\ & + \gamma_{07}GRADE8_i + \gamma_{08}SES_i + \gamma_{09}HISPANIC_i + \gamma_{10}BLACK_i + \gamma_{11}WHITE_i \\ & + \gamma_{12}OTHER_i + \gamma_{13}SPED_i + \gamma_{13}GATE_i + \zeta_{0i}\end{aligned} \tag{2}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}IFEP_i + \gamma_{12}RFEP_i + \gamma_{13}LEP_i + \zeta_{1i} \tag{3}$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}IFEP_i + \gamma_{22}RFEP_i + \gamma_{23}LEP_i \tag{4}$$

$$\pi_{3i} = \gamma_{30} + \gamma_{31}IFEP_i + \gamma_{32}RFEP_i + \gamma_{33}LEP_i \tag{5}$$

where $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_1^2)$, and $\begin{bmatrix}\zeta_{0i}\\ \zeta_{1i}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}\sigma_0^2 & \sigma_{01}\\ \sigma_{10} & \sigma_1^2\end{bmatrix}\right)$.

The coefficient $\gamma_{00}$ represents the average score for EO students at the first wave (the first measurement point); $\gamma_{10}$ represents the average initial slope for EO students; $\gamma_{20}$ represents the average true acceleration for EO students; and $\gamma_{30}$ represents the average summer setback (or gain) for EO students. The random effect $\varepsilon_{ij}$ is a Level 1 residual for student $i$ at time $j$ and is assumed to be drawn from a normal distribution with mean of 0 and variance $\sigma_1^2$. Random effects $\zeta_{0i}$ and $\zeta_{1i}$ represent Level 2 residuals for the intercept and slope, respectively. They are both hypothesized to be drawn from a multivariate normal distribution with a mean of zero, unknown variances $\sigma_0^2$ and $\sigma_1^2$, and unknown covariance $\sigma_{01}$. For each research question, the parameter estimates that are specific to each language proficiency group (i.e., IFEPs: $\gamma_{03}$, $\gamma_{11}$, $\gamma_{21}$, and $\gamma_{31}$; RFEPs: $\gamma_{04}$, $\gamma_{12}$, $\gamma_{22}$, and $\gamma_{32}$; LEPs: $\gamma_{05}$, $\gamma_{13}$, $\gamma_{23}$, and, $\gamma_{33}$) were compared to the reference group parameters (i.e., EO: $\gamma_{00}$, $\gamma_{10}$, $\gamma_{20}$, and $\gamma_{30}$). Parameters $\gamma_{03}$, $\gamma_{04}$, and $\gamma_{05}$ refer to baseline differences in academic vocabulary test scores across language groups with EO as the reference group. Specifically, $\gamma_{03}$ indicates the difference between EO and IFEP students, $\gamma_{04}$ the difference between EO and RFEP students, and $\gamma_{05}$ the difference between EO and LEP students. Similarly, the parameters $\gamma_{11}$, $\gamma_{12}$, and $\gamma_{13}$ refer to differences in rate of growth across the language groups; the parameters $\gamma_{21}$, $\gamma_{22}$, and $\gamma_{23}$ indicate the differences in the acceleration of growth across the language groups, and the parameters $\gamma_{31}$, $\gamma_{32}$, and $\gamma_{33}$ note the differences in the summer setback across language groups. Students' general vocabulary scores from four waves were used as dependent variables and the students' academic vocabulary scores from the first wave (academic vocabulary_W1) were used as a covariate to answer our first research question. The same hypothesized model with academic vocabulary as an outcome variable and general vocabulary scores from the first wave (GV_W1) as a covariate was used to answer our second research question.

Students' vocabulary and reading comprehension scores were used as covariates in our analyses. Specifically, to address our first research question, students' academic vocabulary and reading comprehension scores from the first wave were included in our models as controls. Similarly, students' general vocabulary and reading comprehension scores from the first wave were included in our models as controls to answer our second research questions. We included students' baseline scores as covariates because we wanted to control for students' baseline differences. Vocabulary and reading comprehension skills are strongly correlated, and including these controls allowed us to obtain a more precise estimate of students' initial status in their vocabulary knowledge. These student-level reading-related covariates were not time-varying, and thus do not affect our estimates of students' rate of growth.

The inclusion of the quadratic term improved the model fit in both outcomes (general vocabulary: $\boldsymbol{\Delta}-2\text{LL} = 23.26$, $df = 1$, $p < .001$; academic vocabulary: $\boldsymbol{\Delta}-2\text{LL} = 55.15$, $df = 1$, $p < .001$); its negative value indicates that learning rates decrease as students get older. We ran two separate models with two different dependent variables, which do not allow us to perform direct comparisons across models.

## Results

### Preliminary descriptive analyses

Table 3 displays means and standard deviations of general and academic vocabulary for students of different English proficiency status and grade levels across the four waves. Unsurprisingly, students in older grades performed better than students in lower grades on each measure across language proficiency designations. Comparing performance in the fall of Year 1 and spring of Year 1 reveals that students improved during the school year on both measures in every grade and language category. Trends across summer were much less consistent: some groups improved during the summer while the average scores of other groups dropped. The correlation between general and academic vocabulary measures was .83 in wave 1.

#### *General vocabulary*

IFEP students outperformed all other students across all waves in general vocabulary (Table 3). EO and RFEP students had similar average scores across four waves. Average LEP performance on the general vocabulary knowledge measure was lower than their peers. The rank ordering of average performance across language designations was consistent in all grade level cohorts. All groups

**Table 3** Means and standard deviations of general vocabulary and academic vocabulary scores by students' language designations and grade levels

| | | General vocabulary | | | | Academic vocabulary | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Fall Year 1 | Spring Year 1 | Fall Year 2 | Spring Year 2 | Fall Year 1 | Spring Year 1 | Fall Year 2 | Spring Year 2 |
| 6th grade cohort | EO | 518.08 | 537.33 | 529.34 | 540.97 | −0.15 | 0.29 | 0.27 | 0.37 |
| | | (39.56) | (40.90) | (39.89) | (44.21) | (0.93) | (0.93) | (0.95) | (1.05) |
| | IFEP | 523.03 | 543.44 | 536.71 | 546.13 | 0.10 | 0.43 | 0.60 | 0.70 |
| | | (29.46) | (29.47) | (28.00) | (29.87) | (0.67) | (0.65) | (0.54) | (0.64) |
| | RFEP | 516.92 | 534.87 | 529.95 | 540.25 | −0.10 | 0.29 | 0.40 | 0.61 |
| | | (28.09) | (30.35) | (25.24) | (28.32) | (0.61) | (0.66) | (0.62) | (0.67) |
| | LEP | 480.87 | 493.57 | 494.51 | 500.38 | −0.95 | −0.60 | −0.52 | −0.42 |
| | | (23.82) | (25.61) | (28.17) | (29.97) | (0.64) | (0.59) | (0.72) | (0.79) |
| 7th grade cohort | EO | 525.12 | 538.91 | 541.00 | 550.48 | 0.07 | 0.31 | 0.40 | 0.58 |
| | | (38.18) | (37.79) | (37.09) | (39.74) | (0.95) | (1.04) | (0.92) | (1.00) |
| | IFEP | 537.55 | 544.37 | 552.92 | 560.98 | 0.35 | 0.64 | 0.72 | 0.91 |
| | | (29.23) | (29.86) | (35.30) | (36.85) | (0.78) | (0.78) | (0.75) | (0.78) |
| | RFEP | 526.97 | 537.06 | 541.22 | 552.09 | 0.20 | 0.50 | 0.56 | 0.79 |
| | | (23.30) | (23.82) | (23.44) | (27.02) | (0.62) | (0.62) | (0.63) | (0.63) |
| | LEP | 485.64 | 495.59 | 500.58 | 509.29 | −0.86 | −0.72 | −0.54 | −0.30 |
| | | (27.73) | (29.46) | (27.95) | (30.18) | (0.63) | (0.73) | (0.72) | (0.80) |
| 8th grade cohort | EO | 534.91 | 547.36 | – | – | 0.37 | 0.60 | – | – |
| | | (40.18) | (40.20) | – | – | (0.94) | (1.00) | – | – |
| | IFEP | 545.16 | 558.74 | – | – | 0.60 | 1.01 | – | – |
| | | (31.63) | (31.49) | – | – | (0.80) | (0.76) | – | – |
| | RFEP | 538.62 | 550.32 | – | – | 0.54 | 0.83 | – | – |
| | | (25.62) | (26.90) | – | – | (0.64) | (0.69) | – | – |
| | LEP | 490.67 | 502.33 | – | – | −0.78 | −0.53 | – | – |
| | | (29.37) | (31.00) | – | – | (0.68) | (0.75) | – | – |

Standard deviations in parentheses. Academic Vocabulary scores are scaled scores from two different forms. General Vocabulary scores are the extended scaled scores from Gates-MacGinitie Reading Test. Grade cohorts indicate student's grade levels in the beginning of the study. The descriptives for academic vocabulary reflect the means and standard deviations of the test scores of students who had both general vocabulary and reading comprehension scores at the first wave and other demographic data. Similarly the descriptive statistics for general vocabulary reflect the means and standard deviations of the test scores of students who had both academic vocabulary and reading comprehension scores at the first wave and other demographic data

*EO* English-only, *IFEP* initially fluent English proficient, *RFEP* redesignated fluent English proficient, *LEP* limited English proficient

grew in their general vocabulary across two years: however, some groups (e.g., EO, IFEP, and RFEP students in the sixth grade cohort) experienced a drop in their general vocabulary scores after their summer break. Although students in the seventh grade cohort did not show such pronounced setback, their vocabulary growth during the summer was smaller than it was during the academic year.

*Academic vocabulary*

Across all grade levels, IFEP students outperformed their peers, including EOs, in academic vocabulary across all waves. Given the rigorous redesignation processes used in California, RFEP students had the second highest scores followed by the EO and LEP students. All groups in our sample improved in average academic vocabulary knowledge from wave 1 to wave 4. Among them, IFEPs and RFEPs in the sixth and seventh grade cohort showed the most noticeable improvements. Across groups, summer academic vocabulary growth was slow compared to the pace during the school year.

## Research question 1: What are the general vocabulary trajectories of students in each language designation?

Model A in Table 4 shows students' growth in general vocabulary over two years. There is no difference between the baseline scores of IFEP and EO students (IFEP $\beta = -.63, p = .63$). However, the coefficients for RFEP and LEP students were both negative: in controlled models, RFEP ($\beta = -5.36, p < .001$) and LEP ($\beta = -9.86, p < .001$) students had lower baseline scores than EOs on average. EO students' general vocabulary scores improved by an average of about 2 points per month (TIME $\beta = 1.92$, $p < .001$). Average IFEP and RFEP students' learning rates did not differ from those of EOs controlling for covariates (i.e., there were no TIME by language status interactions for these two groups). However, the interaction of TIME and LEP was negative and statistically significant ($\beta = -.26$, $p = .04$), meaning LEP students' average rate of growth was lower than that of EO students. The negative and significant TIME$^2$ coefficient ($\beta = -.02, p < .001$) indicates that students' rate of growth decreased over time on average. Language group by TIME$^2$ interaction terms were tested but did not improve the model. EO students did experience change in their general vocabulary learning trajectories during the summer (SUMMER; $\beta = -9.42, p < .001$), as did IFEP and RFEP students. LEP students also experienced reduced summer learning, but the setback was not as strong (SUMMER X LFEP $\beta = 5.79, p < .001$). These results are plotted in Fig. 1.

We saw in the descriptives table that IFEP students' had the highest average baseline general vocabulary scores. However, the parameter estimate associated with baseline IFEP students' scores was not significant in our growth model. This apparent difference is most likely explained by the fact that we have good baseline controls in our model, and IFEP students scored higher than their EO peers on the reading comprehension and academic vocabulary covariates. Our prototypical plots demonstrate the predicted growth trajectories of sixth-grade students of each language designation based on the mean scores on the covariates for that group (Fig. 1).[2] They demonstrated that IFEP students (heavy dashed line) began the study with the highest general vocabulary scores,[3] followed by EO (heavy solid line),

[2] This plot is based on the average covariates and parameter estimates of Asian sixth-grade students who is not eligible for free or reduced lunch, not on an individualized education plan, and not in the gifted and talented program.

[3] The results of this multilevel model for change indicate that on average sixth grade EO students scored 524.96 at baseline. This number was calculated by summing the following variables: −377.32

RFEP students (light dashed line), and LEP students (light solid line). Although we plotted the parameterized differences in summer setback from Table 4, these differences were not significant. In essence, the growth of the highest performing groups was the same across the course of the study although there were significant baseline differences. LEP students had flatter rates of growth during the school year and a less pronounced summer setback than students in the other groups.

### Research question 2: What are the academic vocabulary trajectories of students in each language designation?

Model B in Table 4 presents results for EO and LM students' growth in academic vocabulary over two years. The coefficient associated with IFEP was not significant ($\beta = -.01, p = .74$), which indicates that when other variables are controlled, there is no baseline difference between EO and IFEP students. The coefficients associated with RFEP and LEP baseline scores were negative. On average, RFEP ($\beta = -.06$, $p = .01$) and LEP ($\beta = -.21$, $p < .001$) students had lower scores than their EO peers at baseline controlling for demographic information and other assessment scores.

On average, EO students showed growth in academic vocabulary ($\beta = .03$, $p < .001$; .03 points per month). The interaction terms of TIME and language designation (TIME X IFEP, TIME X RFEP, and TIME X LEP) were all positive ($\beta = .01$, $p = .002$, $\beta = .02$, $p < .001$, and $\beta = .01$, $p = .04$, respectively), indicating IFEPs, RFEPs, and LEPs had steeper growth than EOs. The negative and significant TIME$^2$ coefficient ($\beta = -.001$, $p < .001$) indicates that students' rates of growth decreased over time on average. Language group by TIME$^2$ interaction terms were not significant and so were dropped. EO students did not experience summer setback (SUMMER; $\beta = -.01$, $p = .86$) in academic vocabulary. The interaction terms of SUMMER and two language groups (SUMMER X IFEP and SUMMER X LEP) were not significant ($\beta = -.07, p = .22$ and $\beta = .02$, $p = .57$, respectively), indicating neither IFEPs nor LEPs experienced a change in their learning trajectories during the summer. However, the coefficient for SUMMER X RFEP ($\beta = -.09$, $p = .01$) was negative and significant; RFEPs' academic vocabulary learning rates decreased during the summer. These results are illustrated in Fig. 2.

Figure 2 shows prototypical growth trajectories for sixth grade EO, IFEP, RFEP, and LEP students who do not have individualized education plans. At the first wave of data collection, trends were similar to general vocabulary results: IFEP students started the study with the highest average scores, followed by EO, RFEP and LEP students.[4] We note that in our descriptive tables EO students' baseline scores were

Footnote 3 continued
(constant) +20.86 (coefficient for academic vocabulary) ×.004 (mean score of academic vocabulary for EO students at the first wave) +.29 (coefficient for reading comprehension) ×517.22 (mean score of reading comprehension for EO students at the first wave).

[4] The results of this multilevel model for change indicate that on average EO students scored .03 at baseline. This number was computed by summing the following variables: −10.08 (constant) +.01 (coefficient for general vocabulary) ×523.22 (mean score of general vocabulary for EO students at the

lower than RFEP students' scores. This is largely explained by the fact that the descriptives do not control for differences associated with being on an individualized education plan (IEP); a high percentage of EO students had IEPs, producing more divergent predicted values for the (non-IEP) EO prototypical plots than for the other groups. The most obvious trend in Fig. 2 is the relative improvement of IFEP, RFEP, and LEP students relative to EO students across the four waves of data. It is also worth noting that RFEP students lost ground to EO students during the summer, although they more than made up for that during the school year.[5]

## Discussion

The goal of this study was to model the vocabulary growth trajectories of LM and EO students using longitudinal methods. We analyzed the general and academic vocabulary growth trajectories of EO, IFEP, RFEP, and LEP students across middle grades. Students in the study attended schools in a large urban school district in California, a state with rigorous, multiple criteria for redesignating LEP students. At baseline, the rank order of student performance was the same on both vocabulary measures when baseline student-level reading-related skills and characteristics were controlled for. IFEP students performed slightly higher than EO students, who outperformed RFEP and LEP students. However, there were interesting differences in the trajectories of students across the study, which we interpret with reference to both contextual and linguistic factors.

### General vocabulary growth trajectories

We did not find group differences in rates of general vocabulary learning for EO, IFEP, and RFEP students. EO, IFEP, and RFEP students all learned words at a similar rate, and the estimated setback during the summer was also similar. Our results indicate that the IFEPs in our sample outperformed their EO and LM peers in general vocabulary measure. The coefficient associated with the IFEP was not statistically significant when baseline student-level reading-related skills and characteristics were controlled for. However, because IFEPs students had the highest scores in our control variables (academic vocabulary and reading comprehension), their predicted scores were the highest in our sample. Their rate of growth and the amount of summer setback in general vocabulary were similar to those of their EO peers, meaning they were the best performing group on measures of general vocabulary at all time points. We cannot fully explain the reason why

Footnote 4 continued
first wave) +.01 (coefficient for reading comprehension) ×515.84 (mean score of reading comprehension for EO students at the first wave). When IFEP students' mean scores of general vocabulary and reading comprehension were taken into account, the constant for IFEP students was .07. The baseline scores for RFEP and LEP students were calculated to be −.06 and −.88, respectively.

[5] We tested our analytical models with school dummy variables, but inclusion of these variables did not change our results. We also conducted our analysis with special education by time interaction term. However, this interaction was not significant.

**Table 4** Multilevel models for change predicting academic vocabulary and general vocabulary scores

| | General vocabulary Model A | Academic vocabulary Model B |
|---|---|---|
| Fixed effects | | |
| Intercept | 377.32*** | −10.08*** |
| | (7.40) | (0.15) |
| IFEP | −0.63 | −0.01 |
| | (1.30) | (0.03) |
| RFEP | −5.36*** | −0.06* |
| | (0.97) | (0.02) |
| LEP | −9.86*** | −0.21*** |
| | (1.15) | (0.03) |
| TIME | 1.92*** | 0.03*** |
| | (0.10) | (0.002) |
| TIMEXIFEP | −0.05 | 0.01** |
| | (0.16) | (0.004) |
| TIMEXRFEP | 0.05 | 0.02*** |
| | (0.11) | (0.003) |
| TIMEXLEP | −0.26* | 0.01* |
| | (0.13) | (0.003) |
| $TIME^2$ | −0.02*** | −0.001*** |
| | (.004) | (0.0001) |
| SUMMER | −9.42*** | −0.01 |
| | (1.16) | (0.03) |
| SUMMEXIFEP | 2.38 | −0.07 |
| | (2.25) | (0.05) |
| SUMMERXRFEP | 1.70 | −0.09* |
| | (1.50) | (0.04) |
| SUMMERXLEP | 5.79** | 0.02 |
| | (1.74) | (0.04) |
| GV | | 0.01*** |
| | | (0.0003) |
| RC | 0.29*** | 0.01*** |
| | (0.01) | (0.0003) |
| ACV | 20.86*** | |
| | (0.58) | |
| GRADE_7 | −2.57** | −0.10* |
| | (0.74) | (0.02) |
| GRADE_8 | −1.73* | 0.02 |
| | (0.82) | (0.02) |
| WHITE | 4.00** | 0.07* |
| | (1.30) | (0.03) |

**Table 4** continued

| | General vocabulary Model A | Academic vocabulary Model B |
|---|---|---|
| BLACK | −4.07** | −0.10** |
| | (1.29) | (0.03) |
| HISPANIC | −2.32** | −0.01 |
| | (0.78) | (0.02) |
| OTHER | 0.09 | −0.004 |
| | (1.13) | (0.03) |
| SES | −2.49** | −0.04* |
| | (0.84) | (0.03) |
| SPED | −3.47** | −0.17*** |
| | (1.06) | (0.02) |
| GATE | 3.38*** | 0.16*** |
| | (0.74) | (0.02) |
| Level 1 Variance | | |
| Residual | 194.54*** | 0.11*** |
| Component | (4.36) | (0.002) |
| Level 2 Variance | | |
| Intercept | 187.61*** | 0.09*** |
| Component | (9.39) | (0.004) |
| TIME | 0.53*** | 0.0003*** |
| | (.06) | (0.00003) |
| Covariance | −2.63*** | −0.0004 |
| | (.61) | (0.0003) |
| *N* (Students) | 3037 | 3141 |
| *N* (Observations) | 9162 | 9066 |

The reference group in the analysis was English-only students who are in sixth-grade cohort, Asian, and do not receive free or reduced lunch, and do not receive either special or gifted and talented education

*IFEP* initially fluent English proficient, *RFEP* redesignated fluent English proficient, *LEP* limited English proficient, *GV* general vocabulary, *RC* reading comprehension, *ACV* academic vocabulary, *SES* socioeconomic status, *SPED* special education status, *GATE* gifted and talented education status

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

IFEP students in our sample were outperforming their peers, however, it was not unexpected. Previous research has shown that bilingual students can have a *bilingual advantage* in various domains (e.g., Adesope, Lavin, Thompson, & Ungerleider, 2010; Bialystok, 2002, 2011). Cummins' (1979) threshold hypothesis might explain why the IFEPs are high-achievers. This hypothesis suggests that when a bilingual individual attains a certain threshold level of proficiency and competency in his/her first and second language (English in this study context), this individual can enjoy cognitive advantages associated with bilingualism. As IFEP students are, by definition, those who were already fluent English speakers by the time they entered school, they may be fully enjoying their cognitive advantage associated with knowing a second language. On the other hand, kindergarten English proficiency could be a marker of general verbal intelligence, high SES, and high parental investment in education; indeed it is probably a marker of all three. In addition, while the IFEP students may have been aware of two languages that they are exposed to, we could not confirm if the participating IFEP students have any proficiency in their home language. Thus, while these results are consistent with the bilingual advantage hypothesis, they in no way prove it.

On average, EO students showed growth in general vocabulary. In addition, they experienced a summer setback in their general vocabulary. The Fig. 1 indicates that EO students' scores in general vocabulary were lower than their IFEP peers on average and their general vocabulary scores remained higher than their RFEP peers. It is interesting to note that EO, IFEP, and RFEP students showed similar growth in their general vocabulary knowledge. The technical manual for the Gates-MacGinitie Reading Test (our general vocabulary measure) indicates that the words in this test were "words of general usefulness, not obscure or specialized words" (MacGinitie et al., 2000, p. 9). Words used in the test include both high-frequency words and very low-frequency words. The Gates-MacGinitie included words that students are likely to encounter and learn independently, words that are targeted for instruction in school, and also words that are extremely rare and likely only to be encountered from wide independent reading. In that sense, it is not surprising that growth on this measure was stable across groups: the test is intended to sample a wide range of words, and the broader the domain, the more difficult it is for individuals or groups to experience dramatic sustained improvements (Paris, 2005).

Unsurprisingly, the LEP students were the lowest performing group in our study. However, our results indicated that LEP students improved more slowly compared to their EO peers in general vocabulary knowledge during the school year, but continued to learn during the summer. The general vocabulary measure targets some high-frequency words which LEP students might learn independently during the summer, even though they were not able to learn some of the low-frequency words as rapidly as their more proficient peers during the school year. On one hand, it is encouraging to see that LEP students are learning general vocabulary in their middle school years. On the other hand, despite their growth and relative small amount of summer setback, they were indeed the lowest performing students in our sample. LEP students need explicit instruction and attention in promoting their general vocabulary knowledge.

### Academic vocabulary growth trajectories

The results from the multilevel models for change indicate that the IFEP students in our sample outperformed their peers in academic vocabulary measure. In fact, IFEP students showed a steeper rate of academic vocabulary growth than their EO peers. Because the magnitude of summer setback was similar to that of their EO peers, the gap between EO and IFEP students in academic vocabulary was wider at the end of the study than it was in the first wave.

EO students showed growth in academic vocabulary on average. The Fig. 2 indicate that EO students' scores in academic vocabulary were lower than their IFEP peers, and the academic vocabulary scores were predicted to be caught up by the RFEP peers in the second year. It is worth noting that IFEP and RFEP students had steeper academic vocabulary growth trajectories compared to the EO students. It *is* surprising that the IFEP students and RFEP students improved on our academic vocabulary measure more than the EO students given how stable the growth across groups on the general vocabulary measure was. We were worried that estimates of students' vocabulary growth were driven by the differences in their baseline performance, not language status. As a robustness check, we used propensity score matching to find comparable EO and LM students based on their baseline reading, general vocabulary, and academic vocabulary scores, as well as their SES and GATE status. This allowed us to find EO and LM students that have similar reading-related skills and characteristics. Analyses using propensity score matching were consistent with those presented here: EO students did not gain as much in their academic vocabulary as their matched LM peers.

How did IFEP and RFEP students manage to pull ahead in academic vocabulary? It is possible that these IFEP and RFEP students have been instructed to be more strategic in word learning by teachers who emphasize the importance of strategy instruction for second language learners (Oxford, 1994). While it is true that IFEP students do not receive additional English language support once they start schools (e.g., kindergarten or first-grade), we think that IFEPs were likely raised in rich home language environment and/or attended some type of enrichment programs during their preschool years and beyond. LM students may develop greater word awareness in English compared to EO students. When taught academic vocabulary in school, LM students may apply these skills to the new set of challenging words. In contrast, EO students may build a great deal of their general vocabulary implicitly through rich exposure. This could mean that they lack the skills or strategies to learn academic vocabulary as efficiently as their LM peers. In either case, low performing EO students would be good candidates for explicit instruction to word learning strategies and rich exposure to general academic words.

RFEP students had steeper learning trajectories of academic words than EO students. By the end of the study, RFEP students were outperforming the EO students on the academic vocabulary measure. Again, we offer a couple of complementary interpretations. The first is that redesignated students have reached a proficiency threshold in their second language, and that they benefit from a bilingual advantage. Ardasheva, Tretter, and Kinny (2012) used very similar method in their study with redesignated students and found that former LEP students (i.e., RFEPs) outperform their EO and LEP peers in both reading and mathematics. However, such bilingual advantage is not found in all studies, and although the criteria for such bilingual advantage is unclear, there is an easily understood sampling problem in comparing students who achieve high proficiency in two languages with a general sample of monolinguals. Nonetheless, these data including our own might be interpreted as supporting a weak version of the bilingual advantage hypothesis since it was found in academic vocabulary but not in general vocabulary. Another interpretation is that instruction for these students was more explicit or strategic at some point which helped RFEP students to have more strategic word learning capacity than EO students, especially for high-leverage words. Furthermore, redesignated students may have been exposed to more English academic language than LEP students since RFEPs would spend more time in mainstream content-area classroom receiving instruction in content areas.

The LEP students were the lowest performing group in our study in regards to academic vocabulary. Our results indicated that their rate of growth during the school year was more rapid than that of the EO students, however, LEP student in our sample did not experience summer setback in academic vocabulary as their RFEP peers. The academic vocabulary test included no words so infrequent that they would not be encountered in independent summer reading, in an enrichment class, or at a library book club. Given their status as LEP students, more intentional input could have been provided to these students during the summer months. Unfortunately, we did not have any data on their summer activities to back up this argument.

## Other factors influencing students' vocabulary growth trajectories

Linguistic factors are not the only ones that influence student vocabulary growth, and these results from a sample of students attending a large urban district need to also be interpreted with reference to race and SES. The students in our sample were from diverse linguistic and ethnic backgrounds (Tables 1, 2), and ethnicities were not equally distributed across language groups. For instance, while 77 % of the RFEP students were Asians, approximately 30 % of the EO students were Asians. Furthermore, while approximately one quarter of the EO student sample were Black, there were not many Black students in other language groups (0%-2%). Students' racial and ethnic backgrounds are associated with average educational achievement levels in the U.S.: White and Asian students tend to outperform Black and Hispanics (e.g., Farkas & Beron 2004; Fryer & Levitt, 2006; Kao & Thompson, 2003). We tried to control for students' ethnicity by including students' reported ethnicity in all of our analyses. However, statistical controls do not work well when the sample distribution is uneven. In addition, we need to consider demographic factors when interpreting studies in American urban schools which have larger number of LEP students but also larger number of Black students, students from low-income homes, and students who struggle on standardized measure of reading (Dogan et al., 2011).

Table 5 shows means and standard deviations of general and academic vocabulary test scores of White and Black EO students in our study. Unsurprisingly, there was a considerable gap between these two groups of students across two years. Because our main aim was to focus on how students' language status (i.e., whether EO or LM) is associated with their vocabulary growth, investigating the gap between White and Black EO students was out of scope for this paper. However, we acknowledge that Black students who enter school could also have some difficulty acquiring the language of schooling as their fellow LM students as their home language (i.e., African American Vernacular English) also differs from academic English (e.g., Charity, Scarborough, & Griffin, 2004; Terry, Connor, Petscher, & Conlin, 2012). Our EO sample includes a heterogeneous group of students, and treating them as one reference group is likely to have influenced our findings.

**Table 5** Means and standard deviations of vocabulary test scores of White and Black English-only students

| | White | | Black | |
|---|---|---|---|---|
| | M | SD | M | SD |
| General Vocabulary | | | | |
| Fall Year 1 | 553.46 | 37.63 | 504.49 | 34.23 |
| Spring Year 1 | 571.68 | 35.44 | 518.72 | 36.31 |
| Fall Year 2 | 564.82 | 31.89 | 513.43 | 33.08 |
| Spring Year 2 | 579.74 | 38.75 | 524.35 | 38.12 |
| Academic Vocabulary | | | | |
| Fall Year 1 | 0.75 | 0.83 | −0.44 | 0.89 |
| Spring Year 1 | 1.08 | 0.80 | −0.16 | 0.91 |
| Fall Year 2 | 0.95 | 0.77 | −0.22 | 0.85 |
| Spring Year 2 | 1.09 | 0.84 | −0.18 | 1.05 |

Students from all grade-level cohorts were included in this table

The sample for this study comes from a large urban school district that serves a relatively diverse group of students: students from socioeconomically disadvantaged homes, students who are still acquiring English, and students with different ethnicities. This is a distinctive sample from one state in the U.S., and the results that we found from this sample may not generalize to other school districts that serve different group of students. Thus, while these findings provide insight about how students with different language designations may perform over time, we need to be cautious in generalizing these findings across contexts and language groups.

## Limitations and future directions

In interpreting the findings of this study, it is important to consider the demographics of our sample. First, the distribution of first languages in our sample did not match that in the larger U.S. population; 67 % of the LM students in our sample were Asian (53 % being Chinese) and 20 % Hispanic, whereas approximately 75 % of the LM students in the U.S. were from Spanish-speaking homes (National Clearinghouse for English Language Acquisition, 2011). While this may not be a serious limitation of the study, it is important to note that our sample was not representative of LM students nationwide. Additionally, language designation of the participants was solely based on the school district report. We were not involved in the identification and classification of LM students and had to trust the district's decision about students' language proficiency designation. However, the initial performance on the vocabulary measures seemed to be consistent with what we would expect from students in each language designation.

Moreover, while we were able to analyze students' data and find patterns in their vocabulary growth, our explanation of the findings was largely speculative. Unfortunately, we did not have actual data (e.g., students' summer activities, language instruction history) to support our claims. Future studies on LM students'

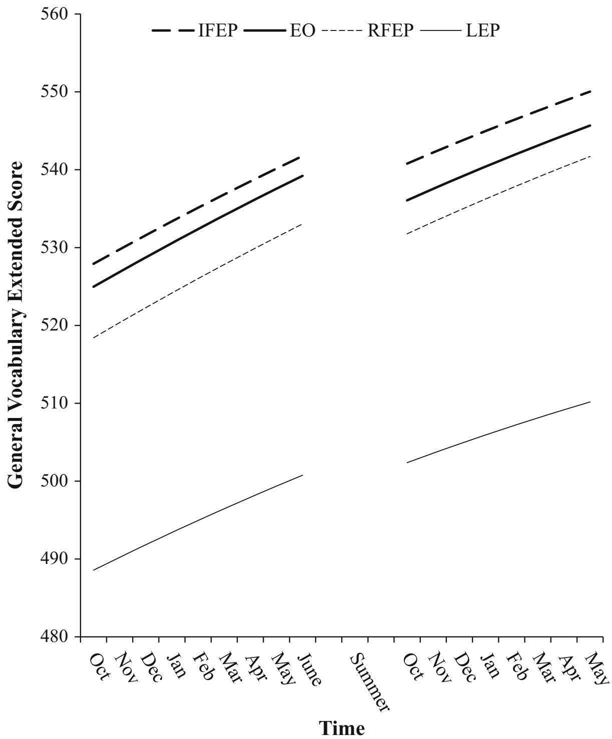


**Fig. 1** Predicted general vocabulary growth trajectories of sixth-grade students of each language designation

vocabulary growth that carefully examine the underlying mechanisms would help us better understand language development of this population.

In regards to assessing students' vocabulary knowledge, we only used one measure for each of the outcomes (i.e., general and academic vocabulary). The two assessments that we used were all synonym tasks which asked the students to find the closest synonym of the target word that was embedded in a sentence (or a phrase) among four to five choices. Using multiple measures to assess different dimensions of word knowledge would have produced more precise assessment of students' vocabulary knowledge. However, our ongoing research into vocabulary assessments using multiple measures suggests that the correlations between student performance across item types are very high, especially if latent scores are used. In addition, we used the same form of Gates-MacGinitie Reading Test across two years. While sixth-grade cohort students took two different forms during the two-year period (i.e., Level 6 in Year 1 and Level 7/9 in Year 2), the seventh-grade cohort students took the same form across four data collection points. Thus, we cannot eliminate the possibility of students' learning of words through multiple testing. However, each data collection point was at least five months apart that it is

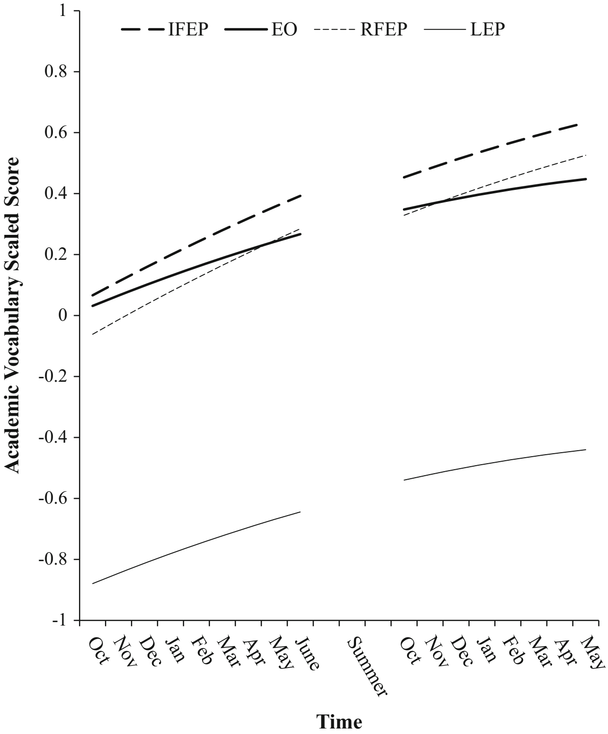


**Fig. 2** Predicted academic vocabulary growth trajectories of sixth-grade students of each language designation

unlikely for students to memorize all the items in the vocabulary subtest. Furthermore, for both general and academic vocabulary tests, students were required to read and decode the sentence stems, distractors, and answers in order to complete the task. However, it was logistically impossible for us to administer our vocabulary assessments orally. There could have been some students who decode poorly in our sample, and our tests are probably not reliable for those students. However, we believed that the vast majority of these students would perform at or below chance on the test and there are very few students in this sample whose decoding and sight word recognition skills are poorer relative to their oral language skills such that inaccurate test results are obtained. Despite these limitations, this paper improves our knowledge of students' general and academic vocabulary knowledge across middle grades and suggests differential mechanisms for word learning.

**Acknowledgments** This research was supported by Grant Number R305A090555, Word Generation: An Efficacy Trial (PI: Catherine Snow) from the Institute of Education Sciences (IES), US Department of Education (USDE).

## References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207–245.

Alexander, K., Entwisle, D., & Olson, L. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*(2), 171–191. doi:10.3102/01623737023002171.

Ardasheva, Y., Tretter, T. R., & Kinny, M. (2012). English language learners and academic achievement: Revisiting the threshold hypothesis. *Language Learning, 62*(3), 769–812. doi:10.1111/j.1467-9922.2011.00652.x.

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., et al. (2011). *The Condition of Education 2011* (*NCES 2011-033*). Washington, DC: U.S. Department of Education, National Center for Education Statistics, U.S. Government Printing Office.

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the national literacy panel on language minority children and youth*. Mahwah, NJ: Erlbaum.

August, D., Carlo, M., Dressler, C., & Snow, C. E. (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.

Bailey, A. L., & Kelly, K. R. (2013). Home language survey practices in the initial identification of English learners in the United States. *Educational Policy, 27*(5), 770–804. doi:10.1177/0895904811432137.

Baumann, J. F., & Graves, M. F. (2010). What is academic vocabulary? *Journal of Adolescent and Adult Literacy, 54*(1), 4–12.

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

Bialystok, E. (2002). Acquisition of literacy in bilingual children: A framework for research. *Language Learning, 52*(1), 159–199.

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology, 65*(4), 229–235.

Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 132–162). Newbury Park, CA: Sage.

Bunch, G. C., Walqui, A., & Pearson, P. D. (2014). Complex text and new common standards in the United States: Pedagogical implications for English learners. *TESOL Quarterly, 48*(3), 533–559. doi:10.1002/tesq.175.

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika, 75*(4), 581–612.

Chall, J. S., & Jacobs, V. A. (2003). Poor children's fourth-grade slump. *American Educator, 27*(1), 14–17.

Charity, A. H., Scarborough, H. S., & Griffin, D. M. (2004). Familiarity with school English in African American children and its relation to early reading achievement. *Child Development, 75*(5), 1340–1356.

Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly, 23*(3), 509–531. doi:10.2307/3586923.

Common Core State Standards Initiative. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.

Corson, D. (1997). The learning and use of academic English words. *Language Learning, 47*(4), 671–718. doi:10.1111/0023-8333.00025.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. doi:10.2307/3587951.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222–251. doi:10.3102/00346543049002222.

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of language and education, Volume 2: Literacy* (2nd ed., pp. 71–83). New York, NY: Springer.

Dogan, E., de Mello, V. B., Lewis, S., Simon, C., Uzzell, R., Horwitz, A., et al. (2011). *Addendum to pieces of the puzzle: Recent performance trends of urban districts—A closer look at 2009 NAEP Results*.

English Language Proficiency Assessment of 1999, California Education Code § § 313–313.5 (2014).

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research, 33*(3), 464–497. doi:10.1016/j.ssresearch.2003.08.001.

Fryer, R. G., & Levitt, S. D. (2006). The Black-White test score gap through third grade. *American Law and Economics Review, 8*(2), 249–281. doi:10.1093/aler/ahl003.

Goldenberg, C. (2010). Reading instruction for English language learners. In M. Kamil, P. D. Pearson, E. Moje, & P. Afflerbach (Eds.), *Handbook of reading research* (Vol. IV, pp. 684–710). Newark, DE: International Reading Association.

Heyns, B. (1978). *Summer learning and the effects of schooling*. New York, NY: Academic Press.

Hiebert, E. H., & Lubliner, S. (2008). The nature, learning, and instruction of general academic vocabulary. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about vocabulary instruction* (pp. 106–129). Newark, DE: International Reading Association.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127–160. doi:10.1007/BF00401799.

Jean, M., & Geva, E. (2009). The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics, 30*(1), 153. doi:10.1017/S0142716408090073.

Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual Review of Sociology, 29*, 417–442. doi:10.1146/annurev.soc.29.010202.100019.

Kim, J., & Herman, J. L. (2009). A three-state study of English learner progress. *Educational Assessment, 14*(3–4), 212–231. doi:10.1080/10627190903422831.

Lawrence, J., Kulesz, P., Hwang, J. K., Francis, D. (under review). Dimensionality of vocabulary knowledge construct.

Lawrence, J. F. (2012). English vocabulary trajectories of students whose parents speak a language other than English: Steep trajectories and sharp summer setback. *Reading and Writing, 25*(5), 1113–1141. doi:10.1007/s11145-011-9305-z.

Lawrence, J. F., Capotosto, L., Branum-Martin, L., White, C., & Snow, C. E. (2012). Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program. *Bilingualism: Language and Cognition, 15*(3), 437–451.doi:10.1017/S1366728911000393.

Leach, J., Scarborough, H., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95*(2), 211–224. doi:10.1037/0022-0663.95.2.211.

Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly, 45*(2), 196–228. doi:10.1598/RRQ.45.2.3.

Lesaux, N. K., Lipka, O., & Siegel, L. S. (2006). Investigating cognitive and linguistic abilities that influence the reading comprehension skills of children from diverse linguistic backgrounds. *Reading and Writing, 19*(1), 99–131. doi:10.1007/s11145-005-4713-6.

Linquanti, R., & Cook, H. G. (2013). *Toward a "Common definition of English learner": A brief defining policy and technical issues and opportunities for State Assessment Consortia*. Washington, DC: Council of Chief State School Officers.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie reading test technical report: Forms S and T*. Chicago, IL: The Riverside Publishing Company.

Mancilla-Martinez, J., & Lesaux, N. K. (2011a). Early home language use and later vocabulary development. *Journal of Educational Psychology, 103*(3), 535–546. doi:10.1037/a0023655.

Mancilla-Martinez, J., & Lesaux, N. K. (2011b). The gap between Spanish speakers' word reading and word knowledge: A longitudinal study. *Child Development, 82*(5), 1544–1560. doi:10.1111/j.1467-8624.2011.01633.x.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47*(1), 91–108. doi:10.1002/RRQ.011.

National Center for Education Statistics. (2012). *The nation's report card: Vocabulary results from 2009 and 2011 NAEP reading assessments (NCES 2013-452)*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

National Center for Education Statistics. (2015). (Table of number and percentage of public school students participating in programs for English language learners, by state: Selected years, 2002-03 through 2012–13). Retrieved from https://nces.ed.gov/programs/digest/d14/tables/dt14_204.20.asp.

National Clearinghouse for English Language Acquisition. (2011). *What langauges do English learners speak? NCELA Fact Sheet*. Retrieved from Washington, DC: http://www.gwu.edu/files/uploads/NCELAFactsheets/EL_Languages_2011.pdf.

National Governors Association. (2010). Common Core state standards for english language arts and literacy in history. In *Social studies, science, and technical subjects-Appendix a: Research supporting key elements of the standards—Common core state standards initiative*. Retrieved from www.Corestandards.Org/the-Standards.

Olsen, L. (2010). *Reparable harm: Fulfilling the unkept promise of educational opportunity for California's long term English learners*. CA: Retrieved from Long Beach.

Oxford, R. L. (1994). *Language learning strategies: An update*. Washington, DC: Eric Clearinghouse on Languages and Linguistics.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184–202. doi:10.1598/RRQ.40.2.3.

Parrish, T., Perez, M., Merickel, A., & Linquanti, R. (2006). *Effects of the implementation of Proposition 227 on the education of English learners, K-12: Findings from a five year evaluation (Final Report)*. Palo Alto and San Francisco, CA: American Institutes for Research and WestEd.

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.

Ragan, A., & Lesaux, N. K. (2006). Federal, state, and district level English language learner program entry and exit requirements: Effects on the education of language minority learners. *Education Policy Analysis Archives, 14*(20), 1–32. doi:10.14507/epaa.v14n20.2006.

Ryan, C. (2013). *Language use in the United States: 2011*. Washington, DC: US Census Bureau, August.

Rydland, V., Aukrust, V., & Fulland, H. (2012). How word decoding, vocabulary and prior topic knowledge predict reading comprehension: A study of language-minority students in Norwegian fifth grade classrooms. *Reading and Writing, 25*(2), 465–482. doi:10.1007/s11145-010-9279-2.

Shin, H. B., & Ortman, J. (2011). *Language projections: 2010 to 2020*. Paper presented at the federal forecasters conference, Washington, DC.

Short, D., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners: A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and even occurrence*. New York, NY: Oxford University Press.

Slama, R. B. (2014). Investigating whether and when English learners are reclassified into mainstream classrooms in the United States: A discrete-time survival analysis. *American Educational Research Journal, 51*(2), 220–252. doi:10.3102/0002831214528277.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). New York, NY: Cambridge University Press.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934. doi:10.1037//0012-1649.38.6.934.

Terry, N. P., Connor, C. M., Petscher, Y., & Conlin, C. R. (2012). Dialect variation and reading: Is change in nonmainstream American English use related to reading achievement in first and second grades? *Journal of Speech, Language and Hearing Research, 55*, 55–69. doi:10.1044/1092-4388(2011/09-0257).

Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal, 112*(3), 497–518. doi:10.1086/663301.

Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly*,. doi:10.1002/rrq.104.

Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation technical manual*. Circle Pines, MN: American Guidance Service.

Woodcock, R. W. (1991). *Woodcock language proficiency battery-revised*. Itasca, IL: Riverside.