Title

Vocabulary and Reading Performances of Redesignated Fluent English Proficient Students

Journal

TESOL Quarterly, 51(4)

ISSN

0039-8322

Authors

Hwang, Jin Kyoung
Lawrence, Joshua Fahey
Collins, Penelope
et al.

Publication Date

2017-12-01

DOI

10.1002/tesq.346

Peer reviewed

tesol QUARTERLY
tesol international association

Vocabulary and Reading Performances of Redesignated Fluent English Proficient Students

JIN KYOUNG HWANG
University of California, Irvine
Irvine, California, United States

JOSHUA FAHEY LAWRENCE
University of California, Irvine
University of Oslo, Oslo, Norway

PENELOPE COLLINS
University of California, Irvine
Irvine, California, United States

CATHERINE SNOW
Harvard Graduate School of Education
Cambridge, Massachusetts, United States

In this article, the researchers examined general vocabulary, academic vocabulary, and reading comprehension growth trajectories of adolescent redesignated fluent English proficient (RFEP) students using individual growth modeling analysis. The sample included 1,226 sixth- to eighth-grade RFEP students from six middle schools in an urban school district in California. Students completed up to four waves of reading-related measures during a 2-year time period. Findings indicate that (a) students' scores on vocabulary and reading assessments were positively correlated with their years since redesignation and (b) students on average showed growth over time on all outcomes and the rate of growth did not differ by their years since redesignation. The results suggest that recently redesignated students may need sustained support to ensure continued progress in their English language development.
doi: 10.1002/tesq.346

Given the large numbers of language minority (LM) students in the United States, it is not surprising that there is a growing body of literature addressing how best to meet their academic needs. There have been research studies about their literacy development (August & Shanahan, 2006) and academic achievement (Collier, 1989; Kieffer,

2008) as well as about the assessments (Abedi, 2002; Bedore et al., 2012; Kieffer, Lesaux, Rivera, & Francis, 2009; Solórzano, 2008), pedagogical knowledge demands (Bunch, 2013; Goldenberg, 2008), instructional environments (Harklau, 1994), and high-quality instructional practices required to support these students (Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006). Despite the improvements in our understanding of how LM students learn and what we can do to teach them better, more research is still needed to understand LM students' learning across different phases of English skill development.

Traditionally, LM students have been treated as two homogeneous groups in the research literature: those classified as limited in English proficiency and those classified as fully English proficient either at school entry or after some exposure to English instruction. Only recently have scholars begun to consider the degree of variability within each of those groups in their English proficiency levels and the implications for their educational trajectories. In particular, one might expect to find heterogeneity within groups of students recently redesignated as fully English proficient, especially because the criteria for reclassification are somewhat ambiguous in some states and extremely variable across states in the United States (Ragan & Lesaux, 2006). Given that there are large numbers of LM students with varying levels of English proficiency, our limited understanding of how these students perform in school settings is concerning. Further, there is a pressing need to understand how LM students' growth in reading skills varies as a function of time since redesignation.

Classification of Redesignated Fluent English Proficient Students

For our purposes, LM students are any school-aged students in the United States who hear or speak a language other than English at home (August & Shanahan, 2006). Nationally this is a large group (approximately 11.2 million; Aud et al., 2011) that varies on many dimensions, including home language, socioeconomic status (SES), and English proficiency. There are no national guidelines for identifying LM students or for classifying them as English proficient or not. Although the specific assessments, proficiency criteria, survey instruments, and language use criteria for (re)classification differ widely across states and school districts, most schools or districts use a variation of the following process: Students enrolling in a school for the first time take a home language survey. Those who report hearing or speaking a language other than English at home are identified as LM students. They are then given an English proficiency screening

assessment. If they meet some minimal locally set criteria, they are classified as initially fluent English proficient (IFEP) students; if not, they are classified as limited English proficient (LEP; also commonly referred to as English language learners [ELLs]). According to federal mandates, ELLs receive additional support for English language development. ELLs are assessed annually until they meet a minimum proficiency criterion or set of criteria, whereupon they are *redesignated* fluent English proficient (RFEP) students. RFEP students no longer receive English language development services in their schools. For instance, California law requires schools to use the following criteria to reclassify ELL students: (1) Results on an English proficiency assessment, (2) teacher evaluation, (3) parental opinion, and (4) comparison to English-proficient students in basic skills that demonstrate ELLs' ability to participate in academic curriculum (English Language Proficiency Assessment of 1999, 2014). Many California districts use the California English Language Development Test (CELDT) to satisfy the first requirement and California Standards Test (CST) for the fourth requirement. With the implementation of the Smarter Balanced Assessments (www.smarterbalanced.org), the CSTs are no longer administered or used for the redesignation process.

Timing of redesignation is an important issue that has not received much attention until recently. Different costs and benefits are associated with reclassification depending on when ELLs exit from English language development services. On the one hand, although more lenient criteria leading to earlier exit from ELL services may restrict RFEP students from receiving supportive language services, it may provide LM students with greater access to the mainstream classroom and higher quality instruction. On the other hand, more rigorous criteria lead to later exit from ELL services, yielding a different set of potential costs and benefits: higher costs for school districts and more limited access for students to the general mainstream curriculum. At the same time, later exit could give ELL students more time and opportunity to develop academic English skills. School districts and policymakers need to walk the line to find the right balance of costs and benefits. As of now, it is difficult to draw definitive conclusions about whether early or late dismissal is better either for individual ELLs or for school districts.

Redesignated Fluent English Proficient Students' Academic Performance

It is difficult to draw strong conclusions about the English proficiency of redesignated students. In some studies, RFEP students

perform as well as or even better than norming groups of English-only (EO) students. For example, in a cross-sectional study, Hwang, Lawrence, Mo, and Snow (2015) showed that adolescent RFEP students' reading comprehension and academic vocabulary scores were comparable to those of EO students and improved with more time after redesignation. Similarly, Ardasheva, Tretter, and Kinny (2012) found that former ELLs (i.e., RFEPs) outperformed ELLs and EO students in both reading and mathematics. In contrast, Slama (2014) found that a large proportion of redesignated students experienced academic difficulty, with approximately one quarter of the sample retained in grade at least once after reclassification. The difference between these two sets of findings might be explained by the stringency of the reclassification criteria in the states where the studies were conducted: California and Kentucky versus Massachusetts.

California has a complex, conjunctive set of criteria for reclassification, whereas Massachusetts relies on only one test, focused on English language proficiency, with no requirement to meet general academic achievement milestones. Umansky and Reardon (2014), studying nine cohorts of Latino students in one California district, found that it took 8 years for 50% of the ELLs to be redesignated and that approximately 25% were never reclassified. In contrast, Slama's (2014) discrete-time survival analysis of over 5,000 ELLs in Massachusetts showed 50% redesignation within 3 years after school entry. Rubio (2014) conducted an analysis very similar to Slama's on data from the New York City Public Schools and estimated 4 years to 50% reclassification, perhaps because there is access to bilingual education in New York, which has been shown to slow progress toward reclassification but generate better long-term outcomes (Umansky & Reardon, 2014). Both Rubio and Slama found faster pathways to redesignation for non-Latinos.

Kim and Herman (2009) compared the academic performance of four groups of students—ELLs, recently redesignated former ELLs, ELLs who had been reclassified more than two academic years earlier, and EOs—across three states. They found achievement differences in ELL and non-ELL students' reading, math, and science assessments across different grades in all three states. However, the magnitude and direction of differences were inconsistent. In one state, RFEP students outperformed current ELLs but underperformed EO students. In another state, RFEP students outperformed even the EO students. These results again suggest that the stringency of reclassification criteria influences how quickly students exit the ELL status and therefore the magnitude and direction of achievement gaps among ELL, former ELL, recently reclassified ELL, and non-ELL students. In states like California, then, where the reclassification process requires several kinds of evidence that students are performing at a high level, it is not surprising

that RFEPs may outscore EOs. Nonetheless, it is likely that RFEPs in California, and elsewhere, show considerable variability in their postredesignation learning trajectories (Hwang et al., 2015; Kieffer, 2008).

Vocabulary Knowledge and Reading Comprehension Skills in Middle School

Vocabulary knowledge and reading comprehension skills are essential for all adolescent students to succeed academically. They are fundamental skills because postprimary students *read to learn* (Chall & Jacobs, 2003) new concepts and ideas across multiple content areas. Weak vocabulary and reading skills can be a critical obstacle in all students' learning in secondary schools as texts become more complex and complicated. Furthermore, for students to perform well in school settings, they need to master the specific register of schooling, academic language (Scarcella, 2003; Schleppegrell, 2004). Academic vocabulary is a critical component of academic language, and it can be classified into general and discipline-specific academic words (Beck, McKeown, & Kucan, 2002; Nagy & Townsend, 2012). General academic words are high-leverage words that appear across multiple subjects, whereas discipline-specific words are closely tied to specific content areas (Beck et al., 2002; Nagy & Townsend, 2012). However, teachers may not recognize the need to focus on general academic words due to their abstract and polysemous natures and because they rarely represent key concepts central to the content areas taught (Snow, 2010).

Although LM students, especially those with limited English proficiency, develop basic skills of reading (e.g., decoding, word reading) at rates relatively comparable to EO students, they tend to lag behind on measures of higher order linguistic skills, such as vocabulary knowledge and language comprehension, compared to the EOs (August, Carlo, Dressler, & Snow, 2005; August & Shanahan, 2006; Lesaux, Lipka, & Siegel, 2006; Mancilla-Martinez & Lesaux, 2011). However, proficient LM students, particularly those designated as IFEP, show performance that is comparable to or even better than their EO peers in these reading-related domains (Hwang et al., 2015; Kieffer, 2011; Lawrence, Capotosto, Branum-Martin, White, & Snow, 2012). In a study done by Hwang and colleagues (2015), even when students' SES (operationalized by students' eligibility for free or reduced lunch) was controlled for, IFEPs outperformed their peers whereas ELLs were falling behind their English-proficient LM and EO peers. Kieffer (2011) also found that IFEP students outperformed their EO peers when SES was controlled for at both the student and school level; however, the

reading growth trajectory of LM students who enter school with initially limited English proficiency converged with those of their EO peers from similar SES backgrounds during their middle school years.[1] It seems likely that the timing of RFEP students' redesignation would relate to their growth in reading-related outcomes in middle school years, but no data on that question are yet available.

Adolescent LM Students' Growth Trajectories in Vocabulary and Reading

There is a small but growing literature on adolescent LM students' growth trajectories in reading-related outcomes (Hwang, Lawrence, & Snow, in press; Kieffer, 2008, 2011; Lawrence, 2012; Lawrence et al., 2012; Mancilla-Martinez, Kieffer, Biancarosa, Christodoulou, & Snow, 2011). One line of results from these studies suggests that EO and LM students' baseline scores in reading predict their later reading scores (Kieffer, 2008, 2011; Mancilla-Martinez et al., 2011; Nakamoto, Lindsey, & Manis, 2007). In other words, students who are below average in earlier grades would continue to be below average in later grades, and those who were above average continue to be above average during their school years. The rate of growth for students at different ability levels does not differ throughout their school year, and their growth trajectories are curvilinear; that is, the rate of growth slowly decreases as students get older. Such findings suggest that it is difficult for students who enter school with below-average reading abilities to catch up to their average or above-average peers.

Another line of findings shows heterogeneity in the vocabulary and reading growth trajectories of EO and LM students based on their language status. For example, Lawrence (2012) found that proficient adolescent LM students showed steeper vocabulary growth during the academic year than their EO peers. However, these LM students also experienced more pronounced summer setback (Alexander, Entwisle, & Olson, 2007) than their EO peers. In other words, the amount of vocabulary knowledge lost during the summer, when the usual school instruction does not occur, was greater for LM students than it was for their EO peers. Alternately, Hwang and colleagues (in press) showed that whereas EO and English proficient LM students experienced summer setback in their general vocabulary knowledge, ELLs' magnitude of summer loss was smaller than that of their counterparts. However, ELLs also showed slower growth during the academic year. A

[1] The LM status and English proficiency in Kieffer (2011) reflect students' initial English language proficiency at kindergarten and did not change over time.

somewhat different pattern was found for students' academic vocabulary knowledge. Although EO, IFEP, and ELL students did not experience change in their academic vocabulary learning trajectories during summer, RFEP students' learning rates were flatter than EO students' learning rates during the summer. However, RFEP students' academic vocabulary learning trajectories during the school year were steeper than those of their EO peers.

Because these studies were conducted in different contexts with heterogeneous samples using different measures, it is difficult to draw definite conclusions about adolescent LM students' growth trajectories in reading-related outcomes. Furthermore, no study to our knowledge has looked at within-group variation in RFEP students' growth during the academic year and the magnitude of summer setback or gain based on their years since redesignation. There are a very small number of studies on differences within the RFEP student category, and this is a serious limitation at this time. Differences in policies, definitions, and criteria across states make it difficult to reconcile the findings suggesting RFEP students continue to struggle (e.g., Slama, 2014), with data suggesting that they are outperforming their peers (e.g., Ardasheva et al., 2012; Hwang et al., 2015). RFEP students are a large and growing group in the United States, but are poorly understood.

The current study builds on the previous cross-sectional analysis on RFEP students from California (Hwang et al., 2015) by examining their vocabulary and reading comprehension growth across the middle grades for 2 years. This study addressed the following research questions:

1. What are the *general vocabulary* growth trajectories of middle school RFEP students at different intervals after redesignation?
2. What are the *academic vocabulary* growth trajectories of middle school RFEP students at different intervals after redesignation?
3. What are the *reading comprehension* growth trajectories of middle school RFEP students at different intervals after redesignation?

METHODS

Sample

Students from six middle schools in a large urban school district in California contributed to the data for the current study. The participating middle schools served students from sixth to eighth grade. Initially, 1,294 RFEP students contributed at least one wave of data

collection. Of these, 59 did not complete the necessary baseline assessments and 17 did not have requisite demographic data. Our final analytical sample consisted of 1,226 students (Table 1). The majority of the students in this sample (90%) were eligible for free or reduced lunch. Our sample consisted of 76% Asian, 17% Hispanic, and 6% other racial/ethnic backgrounds. In the second year of the study, 453 eighth-grade students left the study (graduated from middle school). As can be inferred from Table 1, the sample in this study is not nationally representative and differs from the samples used in most published studies of LM students in the United States and California. The majority of the students were Asian, and many of them were receiving gifted and talented education.

Procedure

To assess students' general vocabulary, academic vocabulary, and reading comprehension, we administered a researcher-developed academic vocabulary test and standardized vocabulary and reading tests. All assessments were administered four times across two consecutive academic years: once at the beginning of the school year (September/October) and once at the end of the school year (May).

Measures

Time. TIME is a level 1 variable indicating the time since the start of the study when students took the assessments. The data were

TABLE 1
Demographic Information on the Participants

	N
Total Students	1,226
Race	
Asian	76%
Hispanic	17%
Other	6%
Eligible for FRL	90%
Special education	2%
Gifted and talented education	58%
Years since redesignation	
Less than 1 year	17%
Less than 2 years	33%
Less than 3 years	19%
More than 3 years	31%

FRL = Free or reduced lunch.

collected in the fall and spring of two consecutive academic 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).

Years since redesignation. The participating school district provided detailed information about RFEP students' date of redesignation, which we used to create a continuous variable YEARS. This variable indicates how many years prior to the initial wave of measurement RFEP students were considered proficient by the district and thus no longer eligible for additional language support services. It does not indicate anything about years in U.S. schools or grade of redesignation. Values on this variable ranged from 0.18 to 4.93 (M = 2.29, SD = 1.34). A histogram of YEARS shows that the distribution is fairly even except for a peak between 1 and 2 years (Figure 1).

General vocabulary. The Gates-MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) is a group-administered assessment that includes vocabulary and passage comprehension subtests. The vocabulary subtest assesses a wide range of general vocabulary knowledge (GEN_VOCAB), so the score from this test was used as an indicator of students' general vocabulary knowledge. The

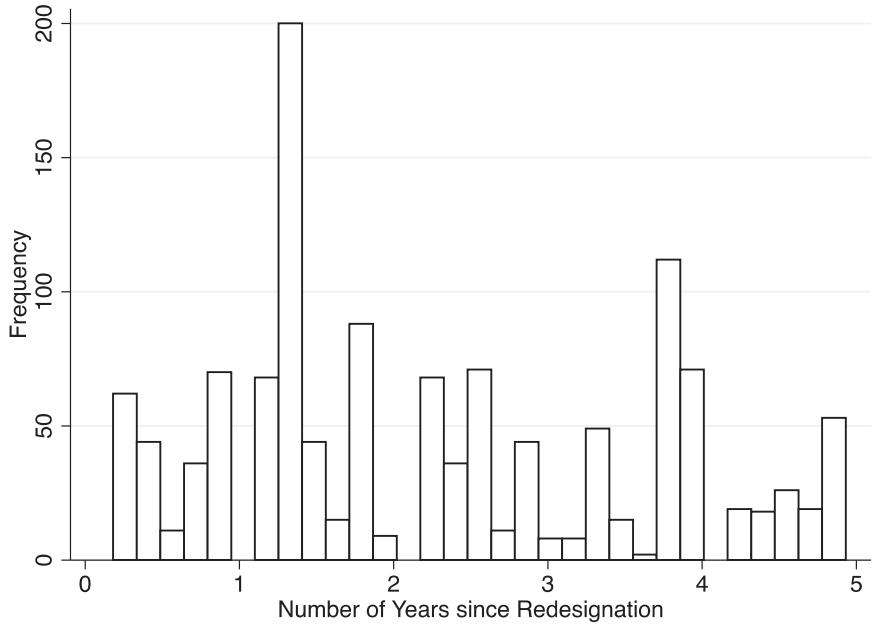

FIGURE 1. Histogram of RFEP students by years since redesignation.

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. 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 were used in the analysis because 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 at seventh grade. The internal reliability (Cronbach's α) for our analytical sample was .84 at the first wave. The mean and the standard deviation of the general vocabulary were 528.59 and 26.89, respectively.

Academic vocabulary. Students' academic vocabulary (ACA_VOCAB) was assessed with a 50-item multiple-choice test that was developed by the research team (all test forms can be found in the Instruments for Research into Second Languages [IRIS] digital repository, www.iris-database.org). For each item, the target word was embedded in a short sentence and students were asked to choose the closest synonym from among four answer choices. The target words were selected to be relatively frequent and dispersed across a range of first-year college texts (Coxhead, 2000).

The same test form was administered twice each year. However, different forms were used in each of the 2 years of the study. The two different forms included 20 anchor items that appeared on both test forms, and the remaining 30 items were unique to each test form. To be able to score responses from the two forms on the same metric, we conducted test scaling using item response theory (IRT) analysis. We ran a unidimensional three-parameter IRT model on the first and third waves of data, constraining the anchor items to have identical item parameters (i.e., difficulty and discrimination). This analysis was done with the sample in the larger study (Hwang et al., in press). In this process, we dropped seven anchor items, two unique items from Year 1 test form, and one unique item from Year 2 test form based on the overall fit index and local dependence statistics. The final model fit the data well (root mean square error of approximation = .03). Marginal reliability for the first wave with the larger sample 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*. These scores were used as an indicator of academic vocabulary in our analysis. The scaled scores had a mean of .25 and standard deviation of .68 at the first wave for our sample.

Reading comprehension. The passage comprehension subtest in the Gates-MacGinitie Reading Test (MacGinitie et al., 2000) measures students' reading comprehension skills (READ). 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 (there are 48 items). Extended scale scores of the reading subtest were used in our analysis. These scores are scaled such that a score of 516 corresponds to average achievement at the beginning of sixth grade and 528 at seventh grade. The Cronbach's α for our analytical sample was .86 in the first wave. The mean and standard deviation of our sample on reading comprehension were 538.01 and 29.27, respectively.

Grade-level cohort. To control for different grade levels in the analysis, student-level dummy variables were created for Grades 6 (GRADE_6), 7 (GRADE_7), and 8 (GRADE_8). There were roughly equal numbers of students from sixth (27%), seventh (37%), and eighth grade (37%) in our analytic sample.

Summer. The number of summers (SUMMER) students had 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 is a time-varying continuous level 1 variable, and its parameter indicates whether students experienced summer setback.

SES. Eligibility for receiving free or reduced lunch 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. Three student-level dummy variables (ASIAN, HISPANIC, and OTHER) were created to control for students' ethnicity. Asian students were used as the reference group in our analysis because they were by far the largest group (76% of the sample).

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 used students' test scores from fall and spring in Year 1 and Year 2 to fit a multilevel model for change describing students' vocabulary and reading skills (Singer & Willett, 2003). The estimates of change in the students' vocabulary and reading skills between data points obtained in this way are more reliable and accurate than those obtained by ordinary least squares regression. This analytical approach was appropriate for our data because we obtained repeated measures (i.e., four data points) of vocabulary and reading for individual students over 2 years. In other words, multiple waves of test data were nested within data for each individual student. The longitudinal models allow us to use all waves of data from each student and obtain robust estimates of growth even with occasional missing data (see the Appendix for the hypothesized multilevel model for change to answer our research questions). Multilevel models have similar assumptions as other general linear models and may provide biased estimates when fit to very small sample sizes (e.g., $N < 50$; Mass & Hox, 2005). However, our analytical sample included more than 1,000 students (level 2), and we are confident in our statistical power to conduct our analyses with these methods.

We calculated the predicted values for seventh-grade RFEP students' general vocabulary, academic vocabulary, and reading comprehension scores based on our final fitted model. We plotted these values as well as the national norms from the Gates-MacGinitie Reading Test (MacGinitie et al., 2000) to compare RFEP students' predicted scores with those of the nationally representative sample.

RESULTS

Preliminary Descriptive Analyses

Table 2 shows correlations among the outcome measures and RFEP students' years since redesignation. The correlation between academic vocabulary and general vocabulary was .79, that between academic vocabulary and reading comprehension was .73, and that between general vocabulary and reading comprehension was .69. As expected, vocabulary and reading measures were highly correlated with one another.

Table 3 displays means and standard deviations of students' general vocabulary, academic vocabulary, and reading comprehension test scores by their years since redesignation across four waves. To clearly

TABLE 2

Correlations Among General Vocabulary, Academic Vocabulary, Reading Comprehension, and RFEP Students' Years Since Redesignation

	General vocabulary	Academic vocabulary	Reading comprehension	Years since redesignation
General vocabulary	1			
Academic vocabulary	0.79	1		
Reading comprehension	0.69	0.73	1	
Years since redesignation	0.47	0.47	0.46	1

illustrate the relationship between RFEP students' years since redesignation and their reading-related outcome scores, we grouped RFEP students into four categories in this table: those who were redesignated less than 1 year before the first assessment, those who were redesignated 1 to 2 years before the first assessment, those who were redesignated 2 to 3 years before the first assessment, and those who were redesignated more than 3 years before the first assessment in this study. Table 3 shows that RFEP students redesignated at different intervals vary in their ability in vocabulary and reading comprehension test scores.

General vocabulary. The first four columns of Table 3 show means and standard deviations for general vocabulary. RFEP students who were redesignated earlier performed better on the general vocabulary measure than those who were redesignated more recently. This pattern was found in all grade-level cohorts. Students gained general vocabulary knowledge over 2 years on average. There was a noticeable summer loss in general vocabulary scores for sixth-grade cohort students (average score at Spring Year 1 = 535.94, average score at Fall Year 2 = 530.35). Although the summer setback was not as pronounced for the seventh-grade cohort students (average score at Spring Year 1 = 536.72, average score at Fall Year 2 = 539.85), the trajectory for vocabulary growth had flattened during the summer.

Academic vocabulary. The fifth through eighth columns of Table 3 describe students' performance on the academic vocabulary measure by years since redesignation. Across all grade-level cohorts, RFEP students' academic vocabulary scores were correlated with postredesignation interval; in other words, the earlier students were redesignated, the higher their baseline test scores. Although students in all groups showed growth in their academic vocabulary over time, the score gap between groups persisted throughout their middle school years. Students did show some growth in their academic vocabulary knowledge

TABLE 3

Means and Standard Deviations of General Vocabulary, Academic Vocabulary, and Reading Comprehension Scores for RFEP Students by Years Since Redesignation

		General vocabulary				Academic vocabulary				Reading comprehension			
	Years since redesignation	Fall Year1	Spring Year1	Fall Year2	Spring Year2	Fall Year1	Spring Year1	Fall Year2	Spring Year2	Fall Year1	Spring Year1	Fall Year2	Spring Year2
Sixth-grade cohort	Less than 1 year	504.79	522.94	520.10	528.17	−0.35	0.02	0.14	0.27	510.79	519.59	533.68	536.85
		(28.56)	(30.05)	(23.83)	(29.42)	(0.57)	(0.66)	(0.60)	(0.66)	(22.31)	(24.48)	(27.65)	(36.73)
	Less than 2 years	518.32	534.98	532.74	543.23	−0.04	0.34	0.50	0.72	523.57	537.88	545.72	555.91
		(25.65)	(27.38)	(24.74)	(25.89)	(0.59)	(0.60)	(0.62)	(0.61)	(22.91)	(25.70)	(26.46)	(30.79)
	Less than 3 years	530.67	549.91	538.20	551.39	0.15	0.56	0.60	0.89	525.48	539.88	547.30	557.78
		(24.46)	(28.69)	(24.00)	(24.96)	(0.59)	(0.65)	(0.55)	(0.60)	(25.83)	(30.23)	(27.65)	(32.49)
	More than 3 years	–	–	–	–	–	–	–	–	–	–	–	–
		–	–	–	–	–	–	–	–	–	–	–	–
Seventh-grade cohort	Less than 1 year	515.86	529.23	532.98	541.66	−0.09	0.18	0.32	0.46	525.68	533.33	541.69	546.14
		(23.23)	(20.91)	(20.48)	(27.15)	(0.61)	(0.59)	(0.57)	(0.62)	(29.21)	(27.93)	(27.71)	(31.22)
	Less than 2 years	522.51	530.97	534.75	546.56	0.14	0.37	0.41	0.69	537.34	545.82	550.51	559.08
		(21.76)	(22.56)	(22.21)	(25.07)	(0.58)	(0.61)	(0.63)	(0.60)	(25.44)	(26.88)	(29.51)	(29.29)
	Less than 3 years	526.03	540.19	539.55	556.68	0.16	0.44	0.69	0.85	538.78	554.65	552.58	555.92
		(23.29)	(22.31)	(18.60)	(27.63)	(0.60)	(0.53)	(0.43)	(0.52)	(21.43)	(25.26)	(21.00)	(36.60)
	More than 3 years	536.85	546.50	552.10	561.12	0.40	0.77	0.81	1.03	547.88	558.29	562.83	571.99
		(21.55)	(23.59)	(22.98)	(26.05)	(0.62)	(0.57)	(0.61)	(0.61)	(27.02)	(29.20)	(32.35)	(27.08)
Eighth-grade cohort	Less than 1 year	518.39	529.29	–	–	0.11	0.38	–	–	527.90	542.30	–	–
		(19.61)	(24.15)	–	–	(0.44)	(0.52)	–	–	(21.53)	(20.61)	–	–
	Less than 2 years	525.72	536.68	–	–	0.17	0.49	–	–	536.29	543.32	–	–
		(22.11)	(21.65)	–	–	(0.57)	(0.61)	–	–	(19.19)	(26.73)	–	–
	Less than 3 years	533.59	544.83	–	–	0.37	0.62	–	–	542.51	553.21	–	–
		(27.34)	(26.25)	–	–	(0.68)	(0.74)	–	–	(29.81)	(28.19)	–	–
	More than 3 years	549.35	561.61	–	–	0.85	1.13	–	–	562.22	570.85	–	–
		(22.00)	(24.39)	–	–	(0.52)	(0.56)			(25.32)	(26.09)	–	–

Note. Standard deviations in parentheses. Grade cohorts indicate students' grade levels at the beginning of the study. Academic vocabulary scores are scaled scores from two different forms of a researcher-developed academic vocabulary test. General vocabulary and reading comprehension scores are the extended scaled scores from the Gates-MacGinitie Reading Test (MacGinitie et al., 2000).

during summer (i.e., between Spring Year 1 and Fall Year 2); however, from these descriptive data it is difficult to tell whether such growth during summer was pronounced enough to change students' overall learning trajectories.

Reading comprehension. RFEP students' reading comprehension test scores were similarly associated with the years since redesignation: Those who had been redesignated more years prior to the assessment scored higher than those who had been recently redesignated. On average, students showed growth in their reading comprehension over the course of 2 years. The amount of gain in students' reading comprehension scores varied according to their years since redesignation. Especially for students in the seventh-grade cohort, those who were redesignated recently showed relatively large gains over the summer break, whereas those who had been redesignated more years prior to the beginning of the study showed relatively small gains or even some setback in their reading comprehension skills. The descriptive results reported above do not allow us to leverage the power of multiple measurement occasions within individuals to improve estimates at each wave nor to control for any covariates. We turn now to longitudinal growth model results to answer each research question.

Growth Modeling Results

Table 4 shows the results from the multilevel models for change predicting general vocabulary, academic vocabulary, and reading comprehension across four waves. The inclusion of the quadratic term ($TIME^2$) improved the model fit for all three outcomes: general vocabulary (Δ2LL = 7.94; $df = 1$, $p < .01$), academic vocabulary (Δ2LL = 18.15; $df = 1$, $p < .001$), and reading comprehension (Δ2LL = 8.54; $df = 1$, $p < .01$). The significant positive linear terms indicate that students tended to improve on these skills across waves. The significant negative quadratic terms indicate that the rates of growth decreased over time.

General vocabulary. Model A in Table 4 presents results for RFEP students' growth in general vocabulary over 2 years. The coefficient associated with YEARS was positive and significant ($\beta = 2.07$, $p < .001$), which indicates that when other variables are controlled, a 1-year interval since redesignation was associated with 2.07 points difference in the baseline score in general vocabulary on average. As is indicated in the technical manual of the Gates-MacGinitie Reading Test (MacGinitie et al., 2000), a score of 515 corresponds to the average general vocabulary score for sixth-grade students at the beginning of their school year and 526 to

TABLE 4

Multilevel Models for Change Predicting General Vocabulary, Academic Vocabulary, and Reading Comprehension Scores

		General vocabulary Model A	Academic vocabulary Model B	Reading comprehension Model C
Fixed effects	Intercept	399.06***	−10.01***	366.15***
		(11.03)	(0.23)	(15.74)
	YEARS	2.07***	0.03**	0.68
		(0.37)	(0.01)	(0.47)
	TIME	1.90***	0.05***	1.56***
		(0.09)	(0.002)	(0.11)
	$TIME^2$	−0.01**	−0.001***	−0.02**
		(0.01)	(0.0001)	(0.01)
	SUMMER	−7.88***	−0.10***	3.24*
		(0.82)	(0.02)	(1.38)
	SUMMER×YEARS			−1.17*
				(0.54)
	GEN_VOCAB		0.01***	0.29***
			(0.001)	(0.03)
	READ	0.24***	0.01***	
		(0.02)	(0.001)	
	ACA_VOCAB	21.24***		15.34***
		(0.89)		(1.23)
	GRADE_7	−4.76***	−0.05*	9.40***
		(1.08)	(0.02)	(1.29)
	GRADE_8	−4.60***	0.07*	12.07***
		(1.28)	(0.03)	(1.52)
	HISPANIC	−3.92***	−0.03	−4.16**
		(1.11)	(0.03)	(1.36)
	OTHER	−1.84	0.04	−0.48
		(1.68)	(0.04)	(2.05)
	FRL	−3.60**	−0.01	1.65
		(1.34)	(0.03)	(1.64)
	SPED	−7.25*	−0.03	−5.64***
		(3.35)	(0.07)	(4.06)
	GATE	0.67***	0.13***	11.42***
		(0.96)	(0.02)	(1.14)
Level 1 variance component	Residual	132.47***	0.09***	174.67
		(4.64)	(0.003)	(6.15)
Level 2 variance component	Intercept	138.77***	0.06***	197.23***
		(10.57)	(0.01)	(14.39)
	TIME	0.44***	0.0002***	0.79***
		(0.06)	(0.00004)	(0.10)
	Covariance	−1.49*	−0.0004	−0.99
		(0.68)	(0.0004)	(0.94)
	N (Students)	1,190	1,218	1,182
	N (Observations)	3,687	3,658	3,661

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

GEN_VOCAB = general vocabulary; READ = reading comprehension; ACA_VOCAB = academic vocabulary; FRL = free or reduced lunch; SPED = special education status; GATE = gifted and talented education status.

Note. The reference group in the analysis was students who were in the sixth-grade cohort, were Asian, did not receive free or reduced lunch, and did not receive either special or gifted and talented education.

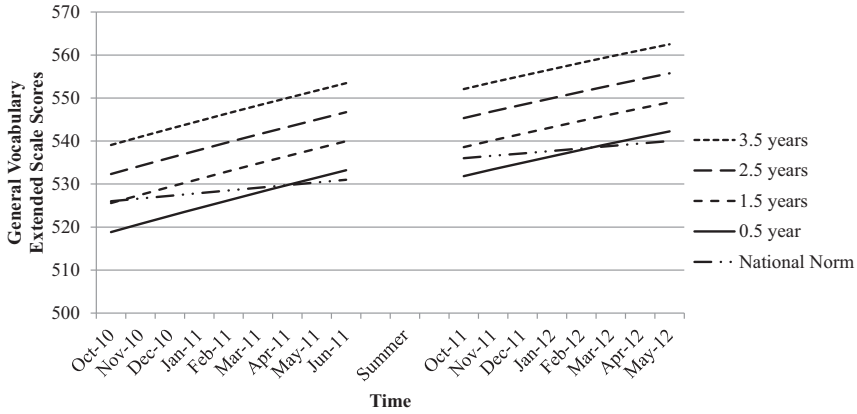

FIGURE 2. Prototypical graph of general vocabulary growth trajectory for RFEP students by years since redesignation.

average score for seventh-grade students. Thus, the expected score gain within one academic year is 11 points (which would be approximately one point per month). With this calculation in mind, the two-point difference could be interpreted as 2 months' worth of growth in students' vocabulary knowledge. On average, RFEP students showed growth in general vocabulary ($\beta = 1.90$, $p < .001$; 1.90 points per month) during this 2-year period. The interaction term of YEARS×TIME was not statistically significant and was not included in our final model. This means that the rate of growth in general vocabulary did not differ by RFEP students' years since redesignation. The negative and significant $TIME^2$ coefficient ($\beta = -0.01$, $p = .005$) indicates that RFEP students' rate of growth decreased over time. Students in our sample experienced summer setback in their general vocabulary knowledge on average (SUMMER; $\beta = -7.88$, $p < .001$). These results are demonstrated in Figure 2.

Figure 2 shows prototypical graphs for seventh-grade RFEP students' general vocabulary growth trajectories. Although YEARS was a continuous variable, we arbitrarily classified our sample into four groups based on their years since redesignation (i.e., 0.5, 1.5, 2.5, and 3.5 years) so that our results are easier to interpret. On average, students who were redesignated 3.5 years before the start of the study (squared dot line, average baseline score = 539.08[2]) already demonstrated significant skill

[2] The average baseline score was calculated by summing the following variables: 399.06 (constant) − 4.76 (coefficient for seventh grade) + 0.24 (coefficient for reading comprehension) × 547.15 (predicted mean score of reading comprehension for seventh-grade RFEP students with 3.5 years since redesignation) + 21.24 (coefficient for academic vocabulary) × 0.41 (predicted mean score of academic vocabulary for seventh-grade RFEP students with 3.5 years since redesignation). The same process for calculating the constant was done for seventh-grade RFEP students with 0.5, 1.5, and 2.5 years since redesignation.

in general vocabulary relative to other RFEP students. Students who were redesignated 2.5 years before the start of the study (long dashed line) did not perform as well (average baseline score = 532.33), but they outperformed students who were redesignated 1.5 years before the start of the study (average baseline score = 525.58) or 0.5 year before the start of the study (average baseline score = 518.83). As we noted in the final fitted model, there is no difference across groups in either growth, acceleration, or summer setback; the lines representing prototypical general vocabulary growth for seventh-grade students are parallel.

Academic vocabulary. Model B in Table 3 presents results for RFEP students' growth in academic vocabulary over 2 years. The coefficient associated with YEARS was positive and significant ($\beta = 0.03$, $p = .003$), which indicates that when other variables are controlled, one year of redesignation was associated with 0.03 point difference in academic vocabulary baseline scores. For instance, the predicted difference in the baseline score of an RFEP student who was redesignated at the beginning of the study (YEARS = 0) and one who was redesignated 1 year before the start of the study (YEARS = 1) would be 0.03 when other variables are controlled for. On average, RFEP students showed growth in academic vocabulary ($\beta = 0.05$, $p < .001$; 0.05 points per month) during their middle school years. Because the academic vocabulary test is a researcher-developed assessment tool and is not normed with a nationally representative sample, we cannot compare 0.03 points to the national average. However, the results from our multilevel models for change indicate that RFEP students in this study are expected to gain 0.05 points per month over this study period on average. Thus, the 0.03-point difference could be interpreted as slightly less than 1 month's growth in academic vocabulary. The interaction term YEARS×TIME was not statistically significant and was not included in our final model. This means that the rate of growth in academic vocabulary did not differ by RFEP students' years since redesignation. The negative and significant $TIME^2$ coefficient ($\beta = -0.001$, $p < .001$) indicates that RFEP students' rate of growth decreased over time. Students in our sample experienced some summer setback on average (SUMMER; $\beta = -0.10$, $p < .001$) in their academic vocabulary knowledge. However, the magnitude of change in their learning trajectory over summer did not differ by years since redesignation.

Figure 3 shows a prototypical graph for seventh-grade RFEP students. On average, students redesignated 3.5 years before the start of the study (squared dot line, average baseline score = 0.30) already demonstrated significantly greater skill in academic vocabulary than other RFEP students. Students who were redesignated 2.5 years before

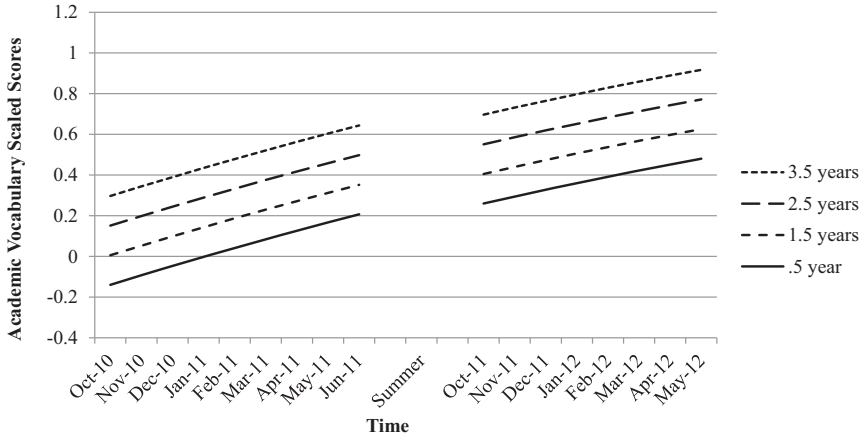

FIGURE 3. Prototypical graph of academic vocabulary growth trajectory for RFEP students by years since redesignation.

the start of the study (long dashed line) did not perform as well (average baseline score = 0.15), but they outperformed students who were redesignated 1.5 years before the start of the study (average baseline score = .01) or 0.5 year before the start of the study (average baseline score = −0.14). As we noted in the final fitted model, there is no difference across groups in growth, acceleration, or summer setback; the lines representing prototypical academic vocabulary growth for seventh-grade students are parallel.

Reading comprehension. Model C in Table 4 presents results for RFEP students' growth in reading comprehension over 2 years. The coefficient associated with YEARS was not statistically significant ($\beta = 0.68$, $p = .145$) when other student-level variables are controlled for. On average, RFEP students showed growth in reading comprehension ($\beta = 1.56$, $p < .001$; 1.56 points per month) during this 2-year period. The interaction of YEARS×TIME was not statistically significant and was not included in our final model. This means that the rate of growth in reading comprehension did not differ by RFEP students' years since redesignation. The negative and significant $TIME^2$ coefficient ($\beta = -0.02$, $p = .003$) indicates that RFEP students' rate of growth decreased over time. The coefficient for SUMMER was positive and statistically significant ($\beta = 3.24$, $p < .001$), which means that, on average, students redesignated within the previous year experienced a gain in their learning trajectories during the summer. However, the coefficient for SUMMER×YEARS was negative and statistically significant ($\beta = -1.17$, $p = .030$). This means that the magnitude of change

in their learning trajectories during the summer differed according to interval since redesignation. These results are demonstrated in Figure 4.

Figure 4 shows a prototypical graph for seventh-grade RFEP students based on our final fitted model. On average, RFEP students who were redesignated 3.5 years before the start of the study (squared dot line, average baseline score = 537.60) demonstrated higher scores in reading comprehension relative to other RFEP students. Students who were redesignated 2.5 years before the start of the study (long dashed line) did not perform as well (average baseline score = 532.75), but they outperformed students who were redesignated 1.5 years before the start of the study (average baseline score = 527.90) or 0.5 year before the start of the study (average baseline score = 523.05). As we noted in the final fitted model, although there was no difference across groups in either growth or acceleration, there was a differential effect of summer producing a slight narrowing of the gaps between groups during their second academic year.

DISCUSSION

We examined RFEP students' vocabulary and reading growth trajectories to understand the potential variability within the RFEP student population. The groups of students showed the same pattern of results for all three outcomes: Those redesignated most recently performed worse, and those redesignated at the greatest interval prior to assessment performed the best. In other words, students who have been

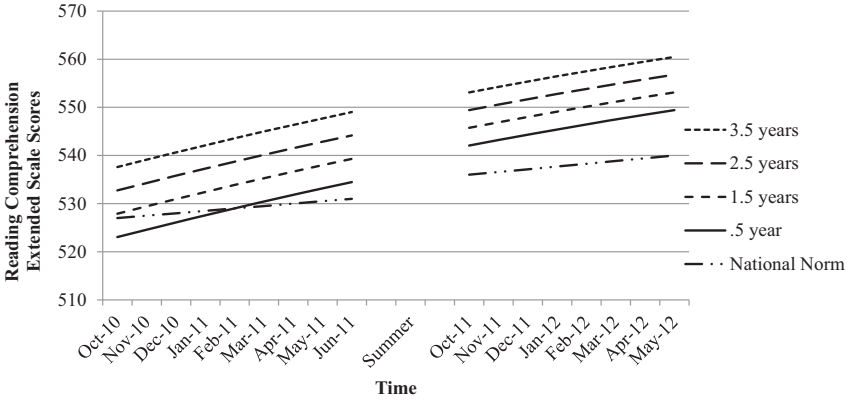

FIGURE 4. Prototypical graph of reading comprehension growth trajectory for RFEP students by years since redesignation.

reclassified as fully English proficient continue to show relative weaknesses in vocabulary and reading comprehension and will require some years to catch up to the levels of students who have been in mainstream classes much longer.

There was no difference in rate of growth over time for any of the three outcomes, and the growth trajectories of the different groups of RFEP students were parallel. RFEP students experienced summer setback for vocabulary and reading outcomes; the summer setback for vocabulary was equivalent across all the RFEP groups, but there was differential effect of summer that slightly closed the gap for reading comprehension skills. We discuss some similarities and differences in RFEP students' vocabulary and reading growth trajectories below.

Similarities in Vocabulary and Reading Comprehension Growth Trajectories

Our results demonstrate that RFEP students' general vocabulary, academic vocabulary, and reading comprehension performance were linearly related to the amount of time since they were redesignated as being proficient in English. That is, although students who were redesignated later in their academic careers may show comparable growth to LM students who had been redesignated in early grades, they did not close the achievement gap in vocabulary and reading comprehension across the middle grades. Our findings, together with past studies, highlight the need to promote proficiency in English early for LM students (Hwang et al., in press; Kieffer, 2008, 2011; Mancilla-Martinez et al., 2011).[3]

The early redesignated students outperformed recently redesignated students in both vocabulary and reading outcomes in the current study. These results are consistent with previous research showing that students' baseline scores predict their later outcomes in reading-related outcomes (Hwang et al., in press; Kieffer, 2008; Mancilla-Martinez et al., 2011; Nakamoto et al., 2007). The rate of growth did not differ among groups of RFEP students at different intervals since redesignation, and the gap among the groups remained throughout their middle school years (especially for vocabulary knowledge). Even so, these rank orders in students' vocabulary and reading scores suggest that recently redesignated students still have room to develop

[3] We are suggesting that early redesignated students have better English literacy outcomes in later grades when they are educated and tested in English. As Umansky and Reardon (2014) found, attending bilingual programs may slow LM students' progress toward reclassification but may generate better long-term academic outcomes.

literacy skills and may benefit from explicit attention to and support for their development of both vocabulary and reading comprehension in order to expedite their learning and enable them to catch up with their early-redesignated peers. With the implementation of the Common Core State Standards, which emphasize the use of academic discourse and higher order language comprehension skills (Common Core State Standards Initiative, 2010), appropriate vocabulary and reading interventions that support and facilitate students' engagement with complex texts and allow them to comprehend and produce academic language would benefit recently redesignated students' academic success in mainstream classrooms (see, for instance, Bunch, Walqui, & Pearson, 2014, for specific pedagogical implications).

An important consideration when interpreting our findings is that California is known to have stringent criteria for redesignation. Indeed, approximately 60% of secondary ELL students in California do not meet the criteria for redesignation after 6 years of instruction (Olsen, 2010; Umansky & Reardon, 2014). Further, the assessment benchmarks used to determine eligibility for redesignation, particularly the state English language arts tests and the CELDT, became more rigorous with increasing grade levels. Thus, it was not surprising to see that our full sample of RFEP students showed high achievement in vocabulary and reading comprehension; even recently redesignated students outperformed the national average eventually. We can anticipate different findings for states that have more lenient criteria for redesignation (e.g., Rubio, 2014; Slama, 2014).

Differences in Vocabulary and Reading Comprehension Growth Trajectories

Our findings highlight the importance of considering specific effects of academic loss over the summer, rather than treating summer setback as a single construct. For example, we found different growth trajectories over the summer for general vocabulary, academic vocabulary, and reading comprehension. For both types of vocabulary, RFEP students' learning trajectories slowed during the summer months. They experienced a loss of approximately 4 months' worth of growth during summer for general vocabulary and approximately 2 months of growth for academic vocabulary on average. Our findings were consistent with others reporting that both EO and (English-proficient) LM students experience summer vocabulary loss in English (Hwang et al., in press; Lawrence, 2012; Lawrence et al., 2012). RFEP students' summer setback in general vocabulary may have been larger than that in

academic vocabulary due to word-level characteristics of target words tested (Paris, 2005). As indicated by the standard frequency index (SFI) in the Educators' Word Frequency Guide (Zeno, Ivens, Millard, & Duvvuri, 1995), the average word frequency of target words in the general vocabulary test was 43.81 (SD = 8.01, range = 22.1–70.3 [higher number indicates higher frequency]). The average SFI for target words in the academic vocabulary test was 48.42 (SD = 5.78, range: 20.8–58.2). Thus, the words that are tested in the general vocabulary test include a wide range of words that vary in their frequency, including very low-frequency words (MacGinitie et al., 2000). It is possible that students' knowledge of very low-frequency words (tested in the general vocabulary assessment) atrophies more sharply. Atrophy may be less sharp for the mid-frequency academic vocabulary words.

We were encouraged by the differential summer reading comprehension trajectories. Students who were redesignated most recently showed greater gains over the summer months than they had during the school year. In contrast, although students who had been redesignated two or more years before the start of the study had higher baseline scores, they showed more pronounced summer setback. Although the differential summer learning reduced the gaps between groups of RFEP students, the underlying mechanism is unclear. It is possible that students who had been redesignated early may not have read as intensively during the summer, thereby experiencing a steeper setback. However, this explanation is speculative because we did not have any data on RFEP students' summer activities to investigate this issue further. In any case, the implication for educators seems to be that summer may be a particularly important leverage point for supporting redesignated language learners.

CONCLUDING REMARKS

Our findings indicate that for RFEP students, number of years since redesignation is positively correlated with vocabulary and reading comprehension outcomes. However, we are not suggesting that early redesignation causes students' higher reading performance. In this study, it was not possible to test the causal relationship between redesignation and students' reading-related performance. In other words, we could not distinguish whether early redesignation helped students benefit from mainstream classroom instruction and develop their language and literacy skills or students who had high linguistic skills were more likely to be redesignated early in their schooling history and were outperforming their peers.

However, we did find that once stringent criteria had been used to redesignate LM students, RFEP students in general performed as well as or even better than students in the national norming samples. Although clear rank order existed among subgroups of RFEP students based on their years since redesignation, the average test scores in reading-related outcomes of the RFEP students in our sample were fairly high. California is known to have rigorous criteria for reclassifying its ELLs. Not only do ELLs need to prove their English proficiency through multiple sources of data (e.g., English proficiency test, content-area standards test, teacher interview, grades in school), they need to meet the cut-off for all the assessments at once. Thus, it is not surprising that RFEP students in this sample performed well on literacy-related assessments. Unfortunately, this high bar also explains why California has a high percentage of long-term ELLs, those who do not get redesignated even after receiving several years of English instruction (Olsen, 2010; Umansky & Reardon, 2014). Thus, rigorous redesignation criteria seem to have two seemingly contrary consequences: high numbers of high-performing RFEP students and high numbers of long-term ELL students.

It is important to consider the demographics of the sample in interpreting our findings. The majority of the students in our sample were Asian, and only 17% were Hispanic, as indicated in Table 1. Generally, students' racial backgrounds were associated with their language and literacy outcomes in the United States, and Asians have been shown to outperform Hispanics and African Americans on average (e.g., Aud et al., 2011; Fryer & Levitt, 2006; Kao & Thompson, 2003). Thus, the distinctiveness of this sample and of the California redesignation procedures require caution in applying current findings to other school districts serving different student populations.

Limitation and Future Directions

One of the shortcomings of this article is that we have no data on students' after-school activities (e.g., debate club, private tutoring) or summer school activities that could have positively influenced their vocabulary and reading outcomes. Future studies that incorporate adolescent RFEP students' extracurricular activities would generate better explanations of their growth in reading-related outcomes. Furthermore, our study did not incorporate any information about curriculum, instruction, or the teacher preparation process in the district where the study took place. Because all the schools that participated in this study were from one school district, we anticipated that their instructional practices would not differ dramatically. Future

studies that carefully examine these issues and how they influence LM students' second language and literacy development would provide a clearer description of the relationship between school instruction and students' language outcomes. Additionally, it is important to acknowledge that the test scores from three measures that were used in this study were a proxy for RFEP students' language and literacy skills. Standardized measures often show low reliability and validity when they are used for LM students (e.g., Abedi, 2002, 2006), and we are not fully satisfied with any measure of student reading comprehension and vocabulary that we have found. However, the measures that we decided to use in this study were adequately reliable. We acknowledge that the scores from these measures do not equate with success in mainstream content area classes; however, we believe there is strong evidence to suggest that they will highly correlate with content-area achievement (e.g., Townsend, Filippini, Collins, & Biancarosa, 2012). Furthermore, as noted earlier, the RFEP students in this study are not a nationally representative sample. The participants of this study were drawn from a single school district, with most students reporting Chinese as their home language and most receiving gifted and talented education, whereas the majority (75%) of the LM students in the United States are from Spanish-speaking homes (National Clearinghouse for English Language Acquisition, 2011). Despite the different student demographics, our findings are convergent with findings involving less advantaged populations (e.g., Kieffer, 2011; Mancilla-Martinez et al., 2011).

Given that the process for redesignation varies to a great degree across states, we need more research on RFEP students to examine whether the current findings hold in other contexts, such as when criteria for redesignation are more lenient. Although more research is needed, and despite the limitations of the current study, these findings fill a gap in the research literature and improve our understanding of adolescent RFEP students' vocabulary and reading development in middle school. The findings of the current study underscore the degree of heterogeneity within the population of adolescent RFEP students; yet despite their different levels of achievement, they show parallel growth trajectories in vocabulary and reading comprehension.

ACKNOWLEDGMENTS

This work was supported by Grant Number R305A090555, Word Generation: An Efficacy Trial, from the Institute of Education Sciences, U.S. Department of Education (Catherine Snow, Principal Investigator).

THE AUTHORS

Jin Kyoung Hwang is a postdoctoral scholar in the School of Education at the University of California, Irvine. Her research centers around understanding the language and literacy development of language minority students, how research-based interventions can help improve their literacy outcomes, and developing and refining test items to accurately assess their language abilities.

Joshua Fahey studies first and second language and literacy development as well as instructional approaches to improve student learning. He is especially interested in experimental program evaluation, longitudinal methods, and vocabulary assessment.

Catherine Snow is the Patricia Albjerg Graham Professor at the Harvard Graduate School of Education. She has studied language and literacy development widely, in monolingual and bilingual learners. She was actively involved in developing the Word Generation program, which is an academic language intervention for middle school students. The Word Generation program has been extended by adding fourth- and fifth-grade units as well as in-depth units focusing on science and social studies curricular topics (see www.wordgen.serpmedia.org).

Penelope Collins is an associate professor in the School of Education at the University of California, Irvine. Her research examines language and literacy development for children from linguistically diverse backgrounds as well as instruction to support academic success for language minority children.

REFERENCES

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment, 8*, 231–257. doi:10.1207/S15326977EA0803_02

Abedi, J. (2006). Language issues in item development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 377–398). Mahwah, NJ: Lawrence Erlbaum.

Alexander, K., Entwisle, D., & Olson, L. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*, 167–180. doi:10.1177/000312240707200202

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

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., & Tahan, K. (2011). *The condition of education 2011 (NCES 2011–033)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

August, D., Carlo, M., Dressler, C., & Snow, C. E. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice, 20*(1), 50–57. doi:10.1111/j.1540-5826.2005.00120.x

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: Lawrence Erlbaum.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford Press.
Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., … Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition*, *15*, 616–629. doi:10.1017/S1366728912000090
Bunch, G. C. (2013). Pedagogical language knowledge preparing mainstream teachers for English learners in the new standards era. *Review of Research in Education*, *37*(1), 298–341. doi:10.3102/0091732X12461772
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*, 533–559. doi:10.1002/tesq.175
Chall, J. S., & Jacobs, V. A. (2003). Poor children's fourth-grade slump. *American Educator*, *27*(1), 14–17.
Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly*, *23*, 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: Author.
Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238. doi:10.2307/3587951
English Language Proficiency Assessment of 1999, California Education Code §§ 313-313.5 (2014).
Francis, D. J., 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. Retrieved from http://www.centeroninstruction.org/files/ELL1-Interventions.pdf
Fryer, R. G., & Levitt, S. D. (2006). The Black-White test score gap through third grade. *American Law and Economics Review*, *8*, 249–281. doi:10.1093/aler/ahl003
Goldenberg, C. (2008). Teaching English language learners: What the research does—and does not—say. *American Educator*, *32*, 7–23, 42–44.
Harklau, L. (1994). ESL versus mainstream classes: Contrasting L2 learning environments. *TESOL Quarterly*, *28*, 241–272. doi:10.2307/3587433
Hwang, J. K., Lawrence, J. F., Mo, E., & Snow, C. E. (2015). Differential effects of a systematic vocabulary intervention on adolescent language minority students with varying levels of English proficiency. *International Journal of Bilingualism*, *19*, 314–332. doi:10.1177/1367006914521698
Hwang, J. K., Lawrence, J. F., & Snow, C. E. (in press). Defying expectations: Vocabulary growth trajectories of high performing language minority students. *Reading and Writing*.
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
Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology*, *100*, 851–868. doi:10.1037/0022-0663.100.4.851
Kieffer, M. J. (2011). Converging trajectories: Reading growth in language minority learners and their classmates, kindergarten to Grade 8. *American Educational Research Journal*, *48*, 1187–1225. doi:10.3102/0002831211419490

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research, 79*, 1168–1201. doi:10.3102/0034654309332490

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. 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*, 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*, 437–451. doi:10.1017/S1366728911000393

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*, 99–131. doi:10.1007/s11145-005-4713-6

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

Mancilla-Martinez, J., Kieffer, M. J., Biancarosa, G., Christodoulou, J. A., & Snow, C. E. (2011). Investigating English reading comprehension growth in adolescent language minority learners: Some insights from the simple view. *Reading and Writing, 24*, 339–354. doi:10.1007/s11145-009-9215-5

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

Mass, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86–92. doi:10.1027/1614-1881.1.3.86

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

Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing, 20*, 691–719. doi:10.1007/s11145-006-9045-7

National Clearinghouse for English Language Acquisition. (2011). *What languages do English learners speak? NCELA Fact Sheet*. Washington, DC: Author.

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

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

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

Rubio, E. V. (2014). *English language learners in an accountability landscape: Examining reclassification, long-term ELLs, language programs, and grade retention* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses. (Order No. 3662579)

Scarcella, R. C. (2003). *Academic English: A conceptual framework*. Berkeley: University of California Linguistic Minority Research Institute.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Lawrence Erlbaum.
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*, 220–252. doi:10.3102/0002831214528277
Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, *328*, 450–452. doi:10.1126/science.1182597
Solórzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research*, *78*, 260–329. doi:10.3102/0034654308317845
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. *Elementary School Journal*, *112*, 497–518. doi:10.1086/66330
Umansky, I. M., & Reardon, S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal*, *51*, 879–912. doi:10.3102/0002831214545110
Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

APPENDIX

Hypothesized Multilevel Model for Change

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

$$\widehat{GEN\_VOCAB} = \pi_{0i} + \pi_{1i} TIME_{ij} + \pi_{2i} TIME_{ij}^2 + \pi_{3i} SUMMER_{ij} + \epsilon_{ij} \quad (1)$$

Level 2 (student level):

$$\begin{aligned}\pi_{0i} =& \gamma_{00} + \gamma_{01} YEARSi + \gamma_{02} READ_i + \gamma_{03} ACA\_VOCAB_i \\ &+ \gamma_{04} GRADE7_i + \gamma_{05} GRADE8_i + \gamma_{06} FRL_i + \gamma_{07} HISPANIC_i \\ &+ \gamma_{08} OTHER_i + \gamma_{09} SPED_i + \gamma_{10} GATE_i + \zeta_{0i}\end{aligned} \quad (2)$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11} YEARS_i + \zeta_{1i}, \quad (3)$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21} YEARS_i \quad (4)$$

$$\pi_{3i} = \gamma_{30} + \gamma_{31} YEARS_i \quad (5)$$

where

$$\epsilon_{i,j} \sim N(0, \sigma_1^2), \text{ and } \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$$

The coefficient γ_{00} represents the average score for redesignated fluent English proficient (RFEP) students who just have been redesignated at the first wave (the first measurement point), γ_{10} represents the average initial slope for RFEP students with 0 year since redesignation, γ_{20} represents the average true acceleration for these students, and γ_{30} represents the average summer setback (or gain) for these students. The random effect ε_{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 σ_1^2. Random effects ζ_{0i} and ζ_{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 0, unknown variances σ_0^2 and σ_1^2, and unknown covariance σ_{01}. For each research question, one unit change in YEARS variable is associated with γ_{01} change in baseline academic vocabulary test scores, γ_{11} change in their rate of growth, γ_{21} change in the acceleration of growth, and γ_{31} change in the summer learning, controlling for all other covariates. We tested our models with the continuous variable YEARS and also with the dummy variables. Including series of dummy variables in the models did not significantly improve the model fit (e.g., Δ2LL = 1.74; $df = 3$, $p = n.s.$ for the model with academic vocabulary as the outcome).

Parameter estimates (γ_{01}, γ_{11}, γ_{21}, and γ_{31}) from our model were used to answer our first research question, controlling for reading comprehension (READ) and academic vocabulary (ACA_VOCAB) at the first wave. To answer our second research question, the same model with academic vocabulary (ACA_VOCAB) as the outcome and general vocabulary (GEN_VOCAB) and reading comprehension (READ) test scores from the first wave as covariates was used. Similarly, a model with reading comprehension (READ) as the outcome and academic vocabulary (ACA_VOCAB) and general vocabulary (GEN_VOCAB) test scores from the first wave as covariates was used to answer our last research question.