American Educational Research Journal

http://aerj.aera.net

Investigating Whether and When English Learners Are Reclassified Into Mainstream Classrooms in the United States: A Discrete-Time Survival Analysis

Rachel B. Slama

Am Educ Res J 2014 51: 220
DOI: 10.3102/0002831214528277

American Educational Research Journal
April 2014, Vol. 51, No. 2, pp. 220–252
DOI: 10.3102/0002831214528277

Investigating Whether and When English Learners Are Reclassified Into Mainstream Classrooms in the United States: A Discrete-Time Survival Analysis

Rachel B. Slama
American Institutes for Research

Using eight waves of longitudinal data on a statewide kindergarten cohort of English learners (ELs), I examined ELs' tenure in language-learning programs and their academic performance following reclassification as fluent English proficient. I employed discrete-time survival analysis to estimate the average time to and grade of reclassification with and without controlling for socioeconomic status and home language. The average EL exited 3 years after school entry or in second grade; however, the odds that a non-Spanish-speaking EL was reclassified were nearly twice that of their Spanish-speaking EL classmates after controlling for income. Despite reclassification in the early elementary grades, large percentages of the kindergarten cohort experienced later academic difficulties and 22% of the sample was retained in grade.

Keywords: English learners, longitudinal studies, reclassification, survival analysis

Improving academic outcomes for English learners (ELs)—students who come from homes where a non-English language is spoken and who need additional academic support in order to access the mainstream curriculum—is a top education policy priority because they are one of the fastest growing yet lowest performing student populations in the nation (Capps, Fix, Murray, Ost, Passel, & Herwantoro, 2005). ELs comprise 10% of the U.S. pre-K-12 student body, and in states with some of the largest EL

Rachel B. Slama is a researcher at the American Institutes for Research, 201 Jones Road, Suite 1, Waltham, MA 02451; e-mail: *rslama@air.org*. Her research interests include the evaluation of federal and state K-12 English learner (EL) policies and, in particular, the use of longitudinal analysis to improve EL accountability policy. This work was funded with generous support from the American Institutes for Research, the Gordon M. Ambach Fellowship, and the State Farm Companies Foundation Fellowship.

populations such as California and Arizona, ELs make up 25% and 15% of school-age students, respectively (National Clearinghouse for English Language Acquisition; Rolstad, Mahoney, & Glass, 2005). Spanish speakers are the largest and most rapidly growing group in the United States, making up nearly 80% of ELs nationwide (Fry & Gonzales, 2008; Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003). All ELs are at increased risk for academic failure because they face the task of simultaneously developing academic English and content-area knowledge (e.g., Scarcella, 2003). However, Spanish-speaking ELs are at particular risk for academic failure because they live disproportionately in poverty (Capps et al., 2005) and are likely to attend "triply segregated" schools with large proportions of other EL students, minorities, and low-income students—risk factors long associated with lagging academic achievement (e.g., Orfield, 2001; Orfield & Lee, 2006; Rios-Aguilar & Gándara, 2012).

Despite a half century of U.S. federal and state legislation spelling out ELs' access to high-quality, adequately funded language-learning programs (*Castañeda v. Pickard*, 1981; *Lau v. Nichols*, 1974) and the ability to access the mainstream curriculum following exit (Williams, 1991), there is growing evidence that ELs do not have access to mainstream curricula, even after many years in a language-learning program. A synthesis of findings from prior studies on ELs' length of time in language-learning programs and those examining academic outcomes following reclassification reveal that substantial numbers of ELs remain in a language-learning program for most of their school years (not including bilingual and dual language programs) and, of those that do exit fairly quickly, many do not fare well academically following exit. Specifically, Hakuta, Goto Butler, and Witt (2000) and Slama (2011) found that many ELs remain in language-learning programs throughout their entire schooling trajectories—exceeding the average time language acquisition scholars estimate it takes for ELs to develop academic language proficiency (e.g., Hakuta et al., 2000). On the other hand, several studies suggest that former ELs may still lag behind mainstream peers academically after exiting language-learning programs in the elementary grades (de Jong, 2004; Gándara et al., 2003; National Center for Education Statistics [NCES], Institute of Education Sciences, U.S. Department of Education, 2009), and these gaps widen over time (de Jong, 2004; National Center for Education Statistics [NCES], Institute of Education Sciences, U.S. Department of Education, 2009).

However, with the exception of a few high-quality survival analyses (Mitchell, Destino, Karam, & Colón-Muñiz, 1999; Parrish et al., 2006), studies of ELs' time to reclassification have been largely unable to provide an unbiased estimate of the time it takes ELs to exit language-learning programs because they have excluded mobile students—the subgroup of students who are at increased risk of academic difficulties (see Reynolds, Chen, & Hebers, 2009, for a review). In the case of mobile ELs, this increased academic risk may indicate a lower probability of reclassification. In addition, most prior studies have been based on cross-sectional data, in which individuals are not followed over time from initial identification as EL until they are reclassified as fluent English proficient (R-FEP). Last, current federal and state monitoring systems only track R-FEP students' progress for 2 years following reclassification (Massachusetts Department of Elementary and Secondary Education [MDESE], 2009; U.S. Department of Education, 2009), making it impossible in many states to examine long-term academic outcomes for reclassified students with large, representative samples.

Thus, in the present study, I employ discrete-time survival analysis to examine the risk of reclassification over time for a kindergarten cohort of ELs and obtain unbiased estimates of time to reclassification for the whole sample, and then for the largest and fastest growing group of ELs in the United States—Spanish-speaking, low-income students. I report my findings both in terms of years in school and in grade. I also report descriptive statistics over time for the study sample on the *Massachusetts Comprehensive Assessment System* (MCAS)—the statewide English language arts (ELA) and mathematics assessments—and students' retention rates. My findings provide information about access to equality of educational outcomes for one of the fastest growing subpopulations nationwide (Capps et al., 2005). In addition, I provide a statewide portrait of student achievement over time 8 years after identification as EL at U.S. school entry.

Theoretical Framework: Longitudinal Analysis as Means of Improving Accountability for EL Achievement

Federal and State Language and Reclassification Policies

Over the past five decades, federal and state litigation, legislation, and language-learning program mandates have aimed to protect the educational rights of ELs (see Rumberger & Tran, 2010, for a review), yet there is little evidence that any one initiative has directly improved these learners' academic outcomes. A series of Supreme Court cases established the right of ELs to access high-quality, adequately funded language-learning programs aligned with their level of academic English (*Castañeda v. Pickard*, 1981; *Lau v. Nichols*, 1974). Furthermore, a federal memorandum (Williams, 1991) stressed that ELs must experience a timely exit from language-learning programs that restrict their access to mainstream curricula and must be able to access the same curriculum as their mainstream peers following reclassification. However, most recently, as part of *Horne v. Flores*—an Arizona-based Supreme Court case that has been open for more than one decade—a federal district court judge ruled on the constitutionality of a language-learning program structure in Arizona that privileges English acquisition over exposure to the core academic curriculum and that separates ELs from mainstream peers for a substantial part of the day (see Rios-Aguilar & Gándara, 2012, for a review). The court ruled that Arizona's program met standards (as reinterpreted by the Supreme Court) of the Equal Educational Opportunities Act (EEOA) of 1974, which requires states "to take appropriate action to overcome language barriers that impede equal participation by its students in its instructional programs" (20 U. S. C. §1703(f)). Nonetheless, Arizona has been scrutinized for prematurely reclassifying ELs and ELs' failure to access mainstream curricula following reclassification (U.S. Department of Education, 2012). Taken together, these findings suggest a need for longitudinal monitoring of EL student progress following reclassification.

At the state level, language policy over the past two decades has prioritized ELs' rapid acquisition of English over home language maintenance (i.e., bilingual programs). In a series of highly politicized campaigns, three states mandated English as the primary language of instruction, moving most students out of bilingual programs and into sheltered (or structured) English immersion (SEI) classrooms. Proposition 227 in California (see Parrish et al., 2006, for a review), Proposition 203 in Arizona (see Rolstad, Mahoney, & Glass, 2005, for a review), and Question 2 in Massachusetts (see Owens, 2010; Tung et al., 2009, for reviews) all resulted in the movement of large numbers of ELs out of bilingual classes and into SEI programs on the grounds that all ELs could quickly develop English with enough meaningful exposure in the content areas combined with explicit language instruction. In all three states, students were expected to develop sufficient English to perform regular class work in English within 1 year (Gándara et al., 2010).

In Massachusetts, the majority of ELs are enrolled in SEI programs until the school or district determines that they have developed sufficient academic English to perform ordinary class work in English without additional support. The goal of these programs is to help students "catch up to their classmates in academic content areas by learning English as quickly as possible" (MDESE, 2009, p. 4). State documents suggest schools and districts rely primarily on the state English language proficiency assessment in addition to a student's performance on the state content-area assessments, grades, teacher observations, class work, and parent input (MDESE, 2009).

Unintended Consequences of Language and Reclassification Policies

Evidence from states and districts suggests that ELs are not receiving the high-quality language-learning services mandated in the Supreme Court cases. For example, Arizona's mandate that ELs receive 4 hours of English language development (ELD) instruction daily—isolated from content-area instruction—has been criticized for resulting in segregation of ELs in SEI programs with other ELs, failing to expose ELs to mainstream instruction, curriculum, and native English-speaking peers (Arias & Faltis, 2012; Gándara et al., 2010; Rios-Aguilar & Gándara, 2012; Rios-Aguilar, González-Canché, & Sabetghadam, 2012) and for failing to improve EL student achievement (Rios-Aguilar et al., 2012). In Massachusetts and California, ELs may remain too long in language-learning programs without developing the requisite skills to succeed in mainstream classrooms: A 5-year longitudinal evaluation of Proposition 227 found that after 10 years in U.S. schools, fewer than half of all ELs were reclassified and even fewer Latino students (Parrish et al., 2006). In Massachusetts, Slama (2011) found that 60% of secondary ELs had attended U.S. schools for 9 or more years without exiting language-learning programs.

Other studies have reported ELs' tenure in language-learning programs that exceeds the average time language-acquisition scholars estimate it takes for ELs to develop academic language proficiency (Batalova, Fix, & Murray, 2007; Hakuta et al., 2000). For many years the prevailing estimates were that it would take 3 to 5 years for the young EL to become proficient in conversational English after entry into an English-speaking environment but 4 to 7 years for these students to develop academic proficiency (Hakuta et al., 2000), and even longer (6 to 8 years) for ELs who immigrated into the country between the ages of 12 and 15 (Collier, 1987). However, more recent estimates suggest that some ELs never catch up to their native peers (Hakuta et al., 2000).

On the other hand, several studies suggest that even ELs who exit language-learning programs in the elementary grades may still lag behind mainstream peers academically after reclassification (de Jong, 2004; Gándara et al., 2003; NCES, Institute of Education Sciences, U.S. Department of Education, 2009). For instance, only 16% and 17% of reclassified ELs scored proficient in eighth-grade reading and mathematics on the NAEP compared to 32% and 35% of monolingual students nationwide (NCES, Institute of Education Sciences, U.S. Department of Education, 2009). However, these studies have largely been based on cross-sectional data (de Jong, 2004; NCES, Institute of Education Sciences, U.S. Department of Education, 2009) and thus do not monitor student performance over time from initial identification as EL to R-FEP status.

While the relationship between time spent in language-learning programs and academic achievement is an important question for future research, the academic difficulties of many early- and late-exit ELs suggest that large proportions of ELs are not faring well academically, regardless of their time spent in language-learning programs.

Reclassification decisions occur within a complex accountability system, which has influenced educators to expedite or prolong ELs' exit from language-learning programs. Under Title I funding streams, schools and districts have an accountability-driven incentive to keep their top-performing ELs classified as limited English proficient in order to inflate this subgroup's performance (Kieffer, Lesaux, & Snow, 2008). Linquanti (2001) described this problem as the "redesignation dilemma." Scholars have long called for an EL designation that follows students from school entry and after reclassification, noting that the present method of defining the EL subgroups leads to inaccurate data, a weakened accountability system, and inappropriate language-learning supports (e.g., Kieffer et al., 2008; The Working Group on EL Policy, 2011). On the other hand, Title III in the United States—through its tie to EL program funding—incentivizes districts to reclassify students as quickly as possible in order to demonstrate that a greater number of their students have reached proficiency. Consequently, schools may be more likely to retain ELs and other low-performing students in the grade (Bowman-Perrot, Herrera, & Murry, 2010). Retention complicates the estimation of time to reclassification because, for a retained student, the number of years in school may not map uniquely onto the grade in which they are reclassified. Thus, in my study, I account for retained students by providing separate estimates of the timing of reclassification in terms of both years in school and by grade. Furthermore, I provide an estimate of the prevalence of grade retention among the students in the cohort who remained in Massachusetts' schools.

The Challenge in Estimating ELs' Tenure in Language-Learning Programs

Previous estimates of the average time it takes for EL students to be reclassified into mainstream classrooms have been largely unable to provide an accurate picture of ELs' reclassification patterns. Widely cited studies have reported sample averages based on cross-sectional data (e.g., Flores, Painter, Harlow-Nash, & Pachon, 2009) or large-scale surveys of the opinions of district administrators (e.g., Kindler, 2002). Even studies that have examined reclassification patterns by following cohorts of students over time (e.g., de Jong, 2004; Gándara & Merino, 1993; Grissom, 2004; New York City Board of Education Division of Assessment and Accountability, 2000) have largely failed to account appropriately for students who may have "disappeared" from the dataset prior to being reclassified into mainstream classrooms (i.e., those students who were "censored"). This is particularly problematic if the censored students are more mobile and have academic profiles that differentiate them systematically from their nonmobile peers, in ways that are associated with a lower probability of reclassification (e.g., see Reynolds et al., 2009, for a review). Several studies did account for censored students by employing survival analyses (e.g., Mitchell et al., 1999; Parrish et al., 2006) and longitudinal analyses (Ramirez, Yuen, & Ramey, 1991); however, two of these studies occurred in a very different policy context, prior to the passage of English-only mandates and only at the district level in California (Mitchell et al., 1999; Ramirez et al., 1991). No studies to date have used these methods in Massachusetts.

In the present study, I employ discrete-time survival analysis to examine the risk of reclassification over time for the first kindergarten cohort of ELs after the passage of Question 2. I obtain unbiased estimates of time to and grade of reclassification and report descriptive statistics of achievement and retention rates. I demonstrate how longitudinal analyses are a promising means of improving accountability for ELs as they allow for monitoring of student outcomes over time, particularly after reclassification as fluent-English proficient. The study is guided by three primary research questions:

Research Question 1: After entering Massachusetts' public schools in kindergarten, what is the average time to first reclassification and the distribution of reclassification over time among ELs?
Research Question 2: In what grade is the average EL in Massachusetts first reclassified, and what is the distribution of reclassification across grades?
Research Question 3: Among students who remained in Massachusetts' public schools, what proportion is retained in the grade during the period of analysis?

In addition to examining the overall time to and grade of first reclassification for all ELs in Massachusetts (RQ1and RQ2, respectively), I examined whether ELs' risk profiles differed for the largest group of ELs in my sample: Spanish-speaking low-income students.

Research Design

Dataset

I constructed a longitudinal dataset comprised of student-level demographics and enrollment from the Massachusetts Student Information Management System (SIMS) and standardized test-performance data from the MCAS. Each student in the state has a unique student identifier that remains with them over time and is common across data sources even if they change schools or districts within the state. This identifier allows me to link student records together across multiple years, regardless of intrastate mobility or retention. My dataset included student-level data for the EL cohort.

Analytic Sample

In the present study, I focus on those ELs who were kindergarteners attending Massachusetts public schools in the fall of 2002. I focus on this cohort because it provides lengthy longitudinal data—eight waves, in total—on the academic careers of these ELs. My sample size is 5,354, which provides sufficient power to detect small effects at the usual levels of Type I error. In Table 1, I show that the sample is predominantly U.S.-born (80.1%), low-income (69.6%), and Spanish-speaking (55.1%), which is similar to nationwide demographic trends with respect to income status yet differs

Table 1

Sample Means for the 2002 Massachusetts Kindergarten Cohort of ELs for the Whole Sample (Second and Third Columns) and the Proportion of Students in Reclassified in Each Group (Fourth and Fifth Columns)

	Whole Sample (n = 5,354)		Proportion of Demographic Group Who Were Reclassified	
Variable	*M*	*n*	*M*	*n*
Reclassified during K-7	0.74	3,961	—	—
Low-income	0.69	3,729	0.73	2,721
Spanish-speaking	0.55	2,948	0.71	2,087
Female	0.48	2,595	0.75	1,955
Immigrant student	0.20	1,073	0.74	790
U.S.-born	0.80	4,290	0.74	3,177
Attends school in an urban district	0.78	4,190	0.72	3,046

Note. $n_{\text{whole sample}}$ = 5,354; $n_{\text{reclassified students}}$ = 3,961. A dash denotes values that are not applicable to the particular cell in question.

with respect to the overall proportion of ELs (7% in Massachusetts compared to 10% nationwide) and the proportion of Spanish-speakers (55% in the present sample compared to 75% nationwide; MDESE, 2012; National Clearinghouse for English Language Acquisition). Spanish speakers are by far the largest language group, followed by speakers of other languages (13%), Portuguese (8.6%), and Vietnamese (5.5%).

Language-learning program. The passage of Question 2 legislation in Fall 2002 (effective in classrooms Fall 2003) closed the majority of bilingual programs in Massachusetts and moved students into SEI programs. In Table 2, I illustrate the effect of the passage of this legislation: Before the new legislation went into effect, nearly half of the kindergarten cohort (n = 2,673) was enrolled in a bilingual program. By the following academic year, the vast majority of these students were either moved out of an EL program entirely (row 2, column 1) or into an SEI program (row 2, column 2). While Table 2 reflects a clear shift in program status as recorded in the state database, it is not clear whether programmatic changes were also immediately affected at the instructional level. Furthermore, as noted in Table 2, there is a dramatic decline in the sample of students in bilingual programs. Thus, I do not include program as a covariate in the present analysis, and the study focuses on the average effect across all programs. Consequently, findings of the average time to and grade of reclassification generalize largely to students enrolled in SEI programs.

Table 2

Participation in Language-Learning Programs, by Year for the 2002 Kindergarten Cohort of ELs

	No EL Program	SEI or ESL	Transitional Bilingual	Two-Way Bilingual	Opt Out
Fall 2002	2.19% (*N* = 117)	27.10% (*N* = 1,451)	49.93% (*N* = 2,673)		20.64% (*N* = 1,105)
Fall 2003	45.29% (*N* = 2,202)	49.03% (*N* = 2,384)	1.79% (*N* = 87)	3.89% (*N* = 189)	*
Fall 2004	47.35% (*N* = 2,185)	45.46% (*N* = 2,098)	1.06% (*N* = 49)	3.66% (*N* = 169)	2.47% (*N* = 114)
Fall 2005	40.29% (*N* = 1,781)	53.38% (*N* = 2,360)	0.95% (*N* = 42)	4.05% (*N* = 179)	1.33% (*N* = 59)
Fall 2006	49.28% (*N* = 2,149)	32.15% (*N* = 1,402)	0.60 (*N* = 26)	3% (*N* = 131)	14.97% (*N* = 653)
Fall 2007	60.34% (*N* = 2,573)	23.87% (*N* = 1,018)	0.52% (*N* = 22)	2.35% (*N* = 100)	12.92% (*N* = 551)
Fall 2008	72.91% (*N* = 3,052)	17.30% (*N* = 724)	—	1.41% (*N* = 59)	8.34% (*N* = 349)
Fall 2009	77.07% (*N* = 3,169)	13.64% (*N* = 561)	—	1.14% (*N* = 47)	8.10% (*N* = 333)

Note. *N* = 5,354 students at baseline (Fall 2002). EL = English learner; SEI = sheltered English immersion; ESL = English as a second language. Asterisk indicates a category that was not available in the given year. Bilingual programs include transitional bilingual programs and two-way bilingual programs in Fall 2002. Consistent with MDESE policy, I do not report data for groups smaller than 10 students. Dash indicates the groups that are too small to report.

Table 3
Most Common School Attendance Profiles for the 2002 Kindergarten EL Cohort

Attendance Pattern	%	*n*
Attended Massachusetts schools at least K-7	70.9	3,794
Left state permanently (i.e., were censored)	22.7	1,213
Left state after kindergarten	6.4	345
Left state after first grade	3.8	204
Left state after second grade	3.5	186
Left state temporarily	6.5	347
Total	100	5354

Note. n = 5,354 students.

Student mobility. In Table 3, I list the most common attendance patterns detected among the kindergarten EL cohort upon entry into Massachusetts schools in 2002. Note that the majority of the students in the present sample (70.9%) remained in Massachusetts public schools for the entire 8 years of study observation (through seventh grade for students who were never retained). However, 22.7% of the sample left the state without reentry into Massachusetts public schools during the period of analysis. These students—the majority of whom left in the early elementary grades (K-2)—are "censored" and thus no longer contribute rows to the person-period dataset once they have left. A smaller proportion of the sample (6.5%) left the state and then reentered Massachusetts public schools again during the 8-year period. Studies that fail to account for student mobility (i.e., a form of censoring) when estimating time to or grade of reclassification will provide biased findings, especially if the academic profiles of mobile students differentiate them systematically from their nonmobile peers, in ways that are associated with a lower probability of reclassification. In my study, censored students are permitted to contribute to the estimated risk of reclassification in each time period, as they remain in the risk set—the group of students who are eligible to be reclassified (i.e., had not previously been reclassified)—for all time periods up until the point of exit.

Students with multiple changes in EL status. In Table 4, I show that the majority of the kindergarten EL cohort were either never reclassified during the 8-year period of observation (i.e., censored; 30.5%) or they were reclassified once (59.5%). However, 9.9% of the sample experienced multiple changes in EL status, indicating that they had been R-FEP, then redesignated as EL subsequently, then rereclassified, and so forth. While interdistrict variability in reclassification criteria might account for multiple changes in EL status (e.g., Ragan & Lesaux, 2006), change in district only accounted for about 23.6% of cases of multiple reclassifications in this dataset. Some additional explanations for multiple changes in EL status—which I am unable to

Table 4
Most Common Reclassification Profiles for the 2002 Kindergarten EL Cohort

Reclassification Pattern	%	*n*
Never reclassified during K-7	30.5	1,632
Reclassified once	59.5	3,187
Multiple changes in EL status		
Two changes in EL status	3.4	184
Three changes in EL status	6.3	337
Four changes in EL status	0.3	14
Total	100%	5,354

Note. n = 5,354 students. The group of students that were never reclassified during K-7 also includes those that were censored.

determine based on my dataset—include active monitoring of ELs, unclear reclassification guidelines, and reporting errors. First, with respect to monitoring, districts are required to monitor student progress for up to 2 years following exit and place children back in a language-learning program if they feel they still need additional support to access the mainstream core curriculum (Title VI, EEOA, No Child Left Behind [NCLB], Title III). Second, unclear reclassification guidelines or changes in who makes reclassification decisions in a particular school or district, especially during the earlier years of the period of analysis, may result in multiple changes of EL status. Reporting errors may also account for multiple changes when a student's EL status is not reported correctly to the state.

Because of the issue of multiple changes in EL status, I focus on the time and grade of students' first reclassification upon school entry in kindergarten.

Measures

I have organized my longitudinal data into a "person-period format" (Singer & Willett, 2003). In the dataset, students contribute one row of data for each occasion of measurement that they are present—that is, one for each of the years that they have attended Massachusetts public schools. The person-period format permits me to record the values of variables that are either "time-varying" (e.g., EL status) or "time invariant" (e.g., home language) for each participant over time.

Outcome

Reclassification into mainstream classrooms. Reclass is a dichotomous time-varying variable that I have coded 1 if the student was reclassified into mainstream classrooms during the particular year in question (0, otherwise). In Massachusetts, reclassification reviews occur typically in the spring of each year, such that any instructional changes are effective the following fall (MDESE, 2009). Students who have not been reclassified by the end of the entire period of observation (8 years) are censored.

Question Predictors

Time in school since kindergarten entry. Time is a continuous time-varying variable that records how many years a student has attended Massachusetts public schools since entry at kindergarten, regardless of grade repetition. Its values range from 1 to 8 because there are only data available on these students through seventh grade.

Grade in school. Grade is a continuous time-varying variable that records the student's present grade in school. Its values can range from 1 (kindergarten) to 8 (seventh grade). A student who has been retained in grade will have the same value for grade for 2 consecutive years—that is, for two consecutive rows in the person-period dataset. Because of the prevalence of retention among ELs, re-representing the passage of time in terms of grade made it possible to estimate grade of reclassification and account for those who have been retained.

Spanish-speaking. Spanish is a dichotomous, time-invariant variable that records whether student's home language was Spanish as reported at school entry in kindergarten. In order to be classified as EL upon entry into Massachusetts public schools, ELs must report speaking a language other than English at home (MDESE, 2003). I examine reclassification profiles disaggregated by this largest language group nationally and in the present sample, controlling for income status.

Covariates

Low-income. Low-income is a dichotomous, time-invariant variable indicating whether the student qualified for free or reduced-price lunch (low-income takes on a value of 1) or not (low-income takes on a value of 0). I examine the effect of income on reclassification profiles alone and then control for its effect in all subsequent discrete-time hazard models examining the main effect of Spanish.

School and district codes. Schoolid and Districtid are integer school and district identification codes, respectively. I included these variables in my statistical models as random effects to account for the natural clustering of the 5,354 ELs in the sample within 483 potentially distinctive schools and 112 districts. Nationwide, ELs tend to be clustered in high-poverty, urban schools (Cosentino de Cohen, Deterding, & Chu Clewell, 2005). Consistent with nationwide trends, in Massachusetts, the majority of ELs are clustered in just five of the largest urban school districts: Boston (23%), Worcester (11%), Lowell (7%), Springfield (5%), and Lawrence (5%; MDESE, 2012). In Figure 1, I present descriptive statistics on the demographic characteristics of the schools in my analytic sample to show that the clustering of ELs in

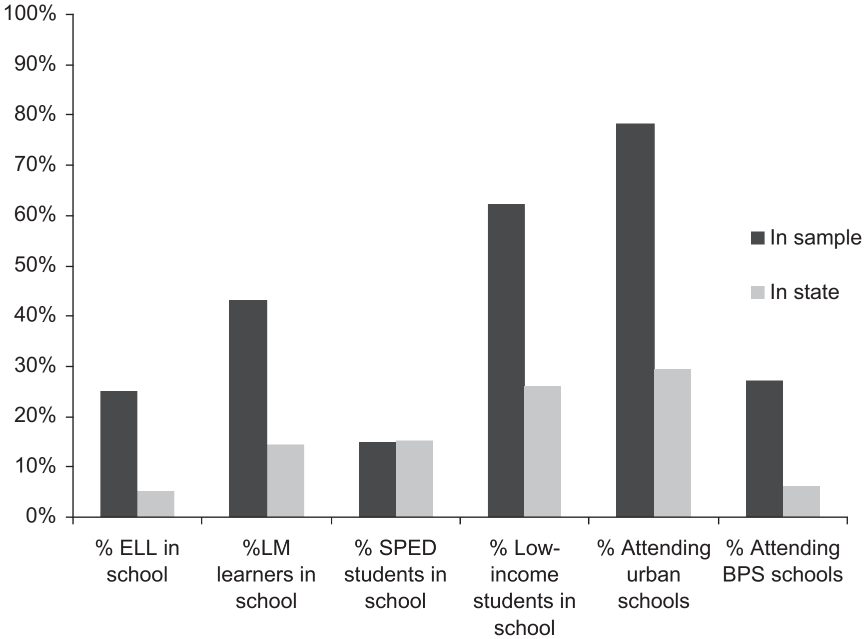

Figure 1. **Average demographic characteristics of sampled schools (*n* = 483 schools; *n* = 112 districts) compared to schools statewide.**

Note. ELL = English language learner; LM = language minority; SPED= special education; BPS = Boston Public Schools.

segregated, high-need schools in Massachusetts replicates the national picture. In Figure 1, I show that the average student in my sample attends a school with large proportions of EL students (25.0% versus 5.2% statewide), language-minority (LM) students (43.2% versus 14.4% in Massachusetts), and low-income students (62.3% versus 26.2% statewide). Furthermore, 78.3% of the students in my sample attend urban schools (compared to 29.4% of students statewide). This segregation of ELs across Massachusetts schools and districts underlines the importance of incorporating school and district random effects into the present analysis.

Data Analysis

To address RQ1 and RQ2, in which I investigate the average time-to-reclassification and grade of reclassification, respectively, I employed a discrete-time survival analysis to calculate the absolute and relative "risk" or probability that a given EL would be reclassified into mainstream classrooms.

All ELs who entered kindergarten in Massachusetts in 2002 were considered part of the risk set—the number of students eligible for reclassification. As the cohort passed through each year or grade of their schooling trajectories, a proportion of them were reclassified into mainstream classrooms, some retained the EL designation, and others left the state or their information was missing from the dataset. Based on this information, in each given year or grade, one can calculate the hazard probability—the proportion of students who began the year as ELs (i.e., the risk set) who were subsequently reclassified during that year or grade. Because hazard probabilities are based on proportions, as students exit language-learning programs or their information becomes missing from the dataset, the pool of students eligible to be reclassified, or the risk set, becomes smaller and the relative "risk" of reclassification can appear greater even if the number of students actually exiting EL programs in a given year or grade is smaller. While this information may be more intuitive conceptually, fitting discrete-time hazard models allows for more complex model specifications and the ability to add covariates and conduct statistical tests. Thus, I am able to examine the risk profiles for the entire kindergarten cohort and then for subgroups at increased risk for academic difficulties; in this case, low-income, Spanish-speaking ELs.

My first task, therefore, was to gain insight into potential model specification. To determine the most appropriate specification of time and grade, I fit a set of parallel taxonomies of discrete-time hazard models by conducting logistic regression analyses in the person-period dataset (Singer & Willett, 2003). First, to estimate the average time to reclassification (RQ1), I fit a baseline discrete-time hazard model specifying the discrete-time hazard probability for EL student i in year j after entry at kindergarten, under the assumption that the given student was not reclassified in the prior (j–1)th year. I then replaced this general specification of time and grade by suitable polynomial specifications of time and grade and compared their fit systematically to lower order polynomial specification as well as to general specification.

I present the results of fitting parallel discrete-time-hazard models of time to and grade of first reclassification and the associated general linear hypothesis [GLH] tests in Tables 5 and 6, respectively. From these comparisons, I determined that the hazard function was best summarized by a cubic specification of time and grade (Model C and Model I, respectively). I systematically included the main effects of Spanish and low-income in the time and grade models, individually (Models D and E in the time taxonomy and Models J and K in the grade taxonomy), together (Models F and L, respectively) and their interaction (not displayed as it was not a significant predictor of risk of reclassification in either set of models; $p > .05$). Based on a series of GLH tests, I conclude that both main effects are required in the time and grade models (Models F and L).

Table 5

Results of Fitting Discrete-Time Hazard Models to the Time to First Reclassification for the 2002–2003 Kindergarten EL Cohort, for the Polynomial Specifications of Time

Variable	Parameter	Model A	Model B	Model C	Model D	Model E	Model F
Intercept	α_0	–0.914*** (0.116)	–1.344*** (0.129)	0.331* (0.164)	0.527*** (0.164)	0.422* (0.165)	0.584*** (0.165)
Time	α_1	0.110*** (0.012)	0.440*** (0.044)	–1.508*** (0.131)	–1.493*** (0.131)	–1.495*** (0.131)	–1.484*** (0.131)
Time^2	α_2		–0.043*** (0.006)	0.528*** (0.037)	0.528*** (0.037)	0.526*** (0.037)	0.526*** (0.037)
Time^3	α_3			–0.047*** (0.003)	–0.047*** (0.003)	–0.047*** (0.003)	–.047*** (0.003)
Spanish	α_4				–0.579*** (0.054)		–0.556*** (0.055)
Low-income	α_5					–0.220*** (0.051)	–0.152* (0.052)
Between-district variance		0.974* (0.101)	0.977* (0.101)	0.947* (0.099)	0.934* (0.098)	0.935* (0.098)	0.931* (0.098)
Between-school variance		0.644* (0.041)	0.639* (0.040)	0.627* (0.040)	0.612* (0.040)	0.624* (0.040)	0.610* (0.040)
Goodness of fit –2LL		17,641.90	17,580.38	17,320.96	17,206.32	17,302.54	17,197.69
Deviance-based hypothesis tests			Compare MA: H_0: $\alpha_2 = 0$ χ^2 (1) = 3.8; $p < .05$; Reject H_0	Compare MB: H_0: $\alpha_3 = 0$ χ^2 (1) = 3.8; $p < .05$; Reject H_0	Compare MC: H_0: $\alpha_4 = 0$ χ^2 (1) = 3.8; $p < .05$; Reject H_0	Compare MC: H_0: $\alpha_5 = 0$ χ^2 (1) = 3.8; $p < .05$; Reject H_0	Compare MC: H_0: $\alpha_4 = 0$; $\alpha_5 = 0$ χ^2 (1) = 3.8; $p < .05$; Reject H_0

Note. $n_{\text{students}} = 5{,}354$; $n_{\text{observations}} = 19{,}189$. M = model; H_0 = null hypothesis. Standard errors in parentheses.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Table 6

Results of Fitting Discrete-Time Hazard Models to the Grade of First Reclassification for the 2002–2003 Kindergarten EL Cohort, for the Polynomial Specification of Grade

Variable	Parameter	Model G	Model H	Model I	Model J	Model K	Model L
Intercept	β_0	–0.969***	–1.339***	0.327***	0.520**	0.427**	0.585**
		(0.115)	(0.129)	(0.168)	(0.168)	(0.169)	(0.169)
Grade	β_1	0.131***	0.424***	–1.542***	–1.533***	–1.53***	–1.522***
		(0.013)	(0.047)	(0.142)	(0.142)	(0.142)	(0.041)
Grade^2	β_2		–0.039***	0.550***	0.551***	0.547***	0.549***
			(0.006)	(0.041)	(0.041)	(0.041)	(.0414)
Grade^3	β_3			–0.050***	–0.050***	–0.050***	–0.050***
				(0.004)	(0.004)	(0.003)	(0.004)
Spanish	β_4				–0.563***		–0.538***
					(0.055)		(0.056)
Low-income	β_5					–0.245***	–0.179**
						(0.052)	(0.053)
Between-district variance		0.951*	0.950*	0.914*	0.899*	0.902*	0.895*
		(0.098)	(0.097)	(0.094)	(0.093)	(0.093)	(0.093)
Between-school variance		0.652*	0.635*	0.062*	0.602*	0.613*	0.601*
		(0.043)	(0.041)	(0.041)	(0.040)	(0.040)	(0.040)
Goodness of fit –2LL		16,543.72	16,501.48	16,272.14	16,168.92	16,250.59	16,157.67
Deviance- based hypothesis tests			Compare MG: H_0: $\beta_2 = 0 \chi^2$ (1) = 3.8; $p < .05$; Reject H_0	Compare MH: H_0: $\beta_3 = 0 \chi^2$ (1) = 3.8; $p < .05$; Reject H_0	Compare MI: H_0: $\beta_4 = 0 \chi^2$ (1) = 3.8; $p > .05$; Reject H_0	Compare MI: H_0: $\beta_5 = 0$ χ^2 (1) = 3.8; $p > .05$; Reject H_0	Compare MI: H_0: $\beta_4 = 0$; $\beta_5 = 0$ χ^2 (1) = 3.8; $p > .05$; Reject H_0

Note. n_{students} = 5,354; $n_{\text{observations}}$ = 19,189. M = model; H_0 = null hypothesis. Standard errors in parentheses.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Table 7
Percentage of Students Scoring at or Above Proficient on the MCAS ELA and Mathematics Assessments, by EL Status

	Whole Sample		Reclassified (R-FEP)		Not Reclassified (EL)		Never-EL	
Grade	ELA	Math	ELA	Math	ELA	Math	ELA	Math
Third	21.3	22.6	60.1	62.7	0.9	1.8	61.0	53.7
Fourth	24.7	24.7	56.0	54.6	1.4	2.9	58.3	49.9
Fifth	25.9	26.7	48.1	48.0	1.8	4.9	63.7	54.1
Sixth	33.0	28.7	55.7	47.2	5.1	6.8	68.7	58.4
Seventh	39.2	29.1	61.7	45.9	13.4	2.8	74.2	54.9

Note. n_{sample} = 5,354; $n_{\text{R-FEP3rd}}$ = 1,857; n_{EL3rd} = 2,598; $n_{\text{R-FEP4th}}$ = 2,314; n_{EL4th} =2,030; $n_{\text{R-FEP5th}}$ = 2,832; n_{EL5th} = 1,388; $n_{\text{R-FEP6th}}$ 3,098; n_{EL6th} = 1,114; $n_{\text{R-FEP7th}}$ = 3,200; n_{EL7th} = 941. MCAS = Massachusetts Comprehensive Assessment System; R-FEP = reclassified as fluent English proficient; ELA = English language arts. MCAS was administered at third grade at which point some students had already left the state.

Based on these model comparisons, I replace the general specification of Time (RQ1) and Grade (RQ2) by cubic specifications in the following respective discrete-time hazard models:

(1) $$\text{Logit hazard } RECLASS_{ij} = \alpha_0 + \alpha_1 GRADE_{ij} + \alpha_2 GRADE^2{}_{ij} + \; + \alpha_3 GRADE^3{}_{ij} + \alpha_4 SPANISH_i + \alpha_5 LOWINCOME_i$$

(2) $$\text{Logit hazard } RECLASS_{ij} = \beta_0 + \beta_1 GRADE_{ij} + \beta_2 GRADE^2{}_{ij} + \; + \beta_3 GRADE^3{}_{ij} + \beta_4 SPANISH_i\,, + \beta_5 LOWINCOME_i$$

where intercept parameters α_0 and $\beta_0$represent the respective values of logit hazard when Time (Equation 1) and Grade (Equation 2) are equal to 0 (as are the values of the covariates). Slope parameters α_1 and β_1 represent the magnitude and direction of change in logit hazard over time. If the signs of these respective parameters are positive, the risk of a given EL being reclassified increases over time; a negative sign indicates that the risk of reclassification declines over time. Parameters α_3 and β_3 signal whether the hazard function first hits a peak or trough and where a positive sign indicates an early peak in the risk of reclassification and a negative sign indicates an early low point in the risk of reclassification and later peak (Singer & Willett, 2003). After fitting these hypothesized models to the data, I was able to recover estimates of all these parameters, construct the corresponding fitted hazard and survivor function, and estimate the median-time to reclassification statewide by years in school (RQ1) and by grade (RQ2). In subsequent models, I added the main effect of time-invariant question predictor, Spanish-speaking to the time and grade models, respectively, controlling for income status. Because Spanish-speaking is a dichotomous variable, taking on a value of 0 for non-Spanish-speaking ELs and 1 for Spanish-speakers, the coefficient's positive sign indicates a greater risk of reclassification among Spanish-speaking ELs and a negative sign indicates a lower risk of reclassification for this group compared to non-Spanish-speaking EL peers. The coefficient's magnitude quantifies the difference in risk of reclassification between the two groups.

Results

Research Questions 1 and 2: Majority of ELs Are Reclassified Within Three Years

To answer Research Question 1 and Research Question 2—investigating the average time to first reclassification and the distribution of reclassification over time, respectively—I summarize evidence from fitted discrete-time hazard models and their accompanying plots. I found that while a majority of ELs who entered at kindergarten were first reclassified within 3 years (or by second grade after accounting for retention), the greatest risk of first reclassification occurred 6 years after school entry (or at the end of fifth grade). While this may seem a little incongruous, as noted earlier, this occurs because the risk set—or the students who are still classified as EL at the beginning of a given school year—in Year 6 is smaller than the risk set in Year 3. Thus, while the absolute number of students who were first reclassified after 6 years in school is smaller, they comprise a greater percentage of the risk set.

The fitted discrete-time hazard models for time to first reclassification (RQ1) and grade of first reclassification (RQ2), presented in Tables 6 and 7, respectively, provides insight into the shape of the hazard functions in the top and bottom left panels of Figure 2. First, in each of the models of risk of reclassification among the entire kindergarten cohort—Models C and I—each of which adopts a cubic specification of time—notice the negative sign on parameters α_1 and β_1. This indicates that the risk of reclassification declines generically over time. A negative sign on parameters α_3 and β_3—which signal whether the hazard function first hits a peak or trough—-indicates an early low point in the risk of reclassification and a later peak. More precisely, from these final models, I estimated that the troughs or lowest points in the risk of reclassification occur after 2.25 years or midway through second grade and the peak risk of reclassification occurs 5.9 years after school entry or at the end of fifth grade (Singer & Willett, 2003).

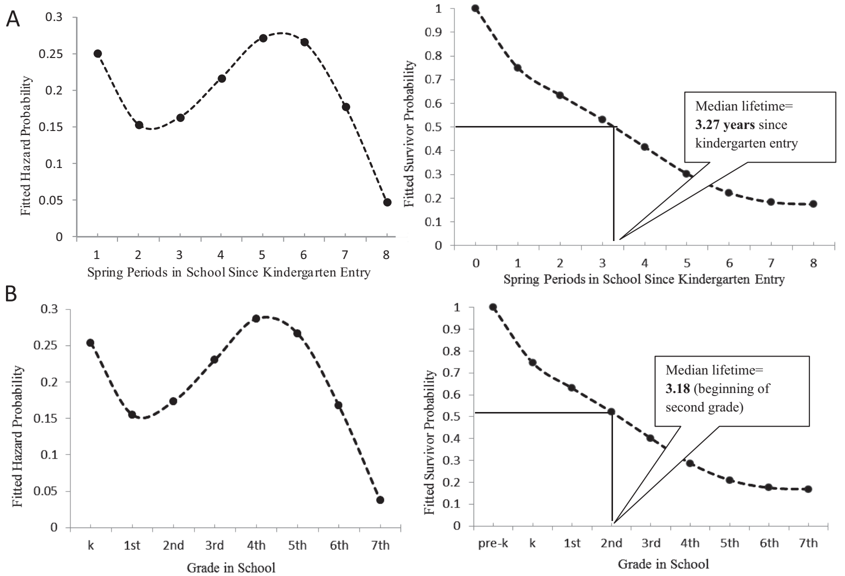

Figure 2. **Estimated hazard (top and bottom left panels) and survivor functions (top and bottom right panels) for the 5,354 ELs in the 2002 kindergarten cohort, by years in school (Panel A; from fitted Model C) and grade in school (Panel B; from fitted Model I), with median lifetimes overlaid on plot of survivor functions.**

To elucidate the complex changing nature of risk over time, I present plots of the estimated discrete-time hazard function over time derived from final Model C and final Model I, respectively, and then map these findings onto their corresponding parameters in a taxonomy of fitted discrete-time hazard models. In top left and bottom left panels of Figure 2, for instance, I illustrate that the overall shape of the hazard functions for the risk of reclassification by Time (Panel A) and Grade (Panel B) decreases over the period of study. Examining this pair of figures, note that the risk of reclassification is lowest after 2 years in school, or at the end of first grade. Further, the peak risk in reclassification occurs after 6 years in school or at the end of fifth grade before declining dramatically over the seventh and eighth year in school (sixth and seventh grade, respectively). Nonetheless, overall risk of reclassification appears to decline over time and across grades. In the top right and bottom right panels of Figure 2, I present the corresponding fitted survivor functions to show that the estimated survivor probability drops to 0.5, or the median life time (Singer & Willett, 2003) occurs after approximately 3.27 years since kindergarten entry (right figure, Panel A) or at the beginning of second grade after accounting for retention (right figure, Panel B). In other words, 50% of the kindergarten EL cohort was reclassified into mainstream classrooms after just over 3 years in U.S. schools or based on the estimate that accounts for retention, following the spring of first grade.

Research Questions 1 and 2: Spanish-Speaking, Low-Income Students Remain Classified as EL

To extend Research Questions 1 and 2, I investigated the average time to and grade of first reclassification for Spanish-speaking ELs in the kindergarten cohort, and found that the odds that a non-Spanish-speaking EL is reclassified is nearly twice that of a Spanish-speaking EL, and the effect persists after controlling for income status. Specifically, in Models F (Table 5) and L (Table 6), which includes the main effects of Spanish and low-income, notice that the coefficient on the parameter associated with Spanish-speaker is negative, indicating that, controlling for income status, the risk of reclassification is lower for Spanish-speakers compared to ELs from other language groups (i.e., Spanish-speaking students remain classified as EL for longer periods of time).

In Figure 3, I graphically display the relationship between time to reclassification, language group, and income status. In Panel A, I demonstrate the effect of income (controlling for whether the student was a Spanish speaker) on the fitted survivor functions, and Panel B displays the effect of Spanish (controlling for income status) on the fitted survivor functions. Specifically, in Panel A, I demonstrate that the average time to first reclassification or median lifetime is just over 4 years after kindergarten entry for a low-income, Spanish-speaking EL (corresponding to the beginning of third grade after accounting for retention) compared to approximately 3 years (beginning of second grade after accounting for retention) for their low-income EL peers from other language groups (Figure 3, Panel B) and classmates not from low-income families but from the same language group (Figure 3, Panel A). In Figure 4, I display the analogous relationship for grade of reclassification, and we observe the same patterns evident in the time to reclassification dataset.

Research Questions 1 and 2: Large Proportions of Reclassified Students Still Struggle With Grade-Level Academic Language

Examining student performance on the state-mandated mathematics and ELA MCAS tests at different time points along their school trajectories indicates that large percentages of ELs in my sample are struggling to keep up with mainstream class work even several years after reclassification into those classrooms. In Table 7, I show that by third grade—when the majority of students in my sample had been reclassified into mainstream classrooms—substantial numbers of students were still scoring below proficient in both ELA and mathematics. Specifically, 60.1% and 62.7% of reclassified ELs scored proficient on ELA and mathematics, respectively, compared to

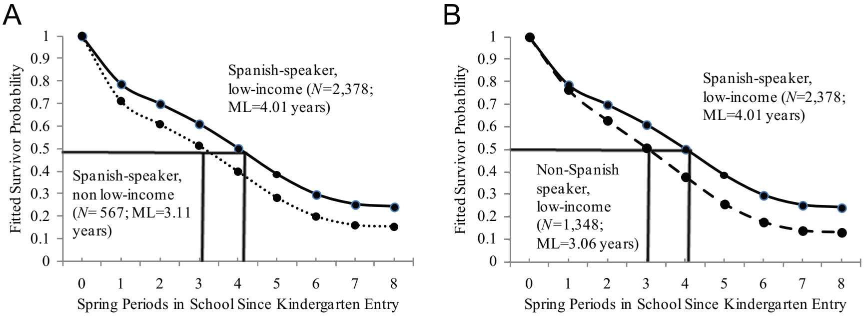

Figure 3. **Estimated survivor functions for prototypical English learner (EL) in the 2002 kindergarten cohort, by years in school, by income status (Panel A), and by whether a student is a Spanish speaker (Panel B) from fitted Model F.**

Note. Median lifetimes (MLs) overlaid on plot of survivor functions.

less than 1% and 2% of the sample who remained classified as EL and 61% and 53.7% of students who were never classified as EL. At the spring of fifth grade—the peak of the conditional probability of reclassification— fewer than half of reclassified students were proficient in mathematics (48%) and ELA (48.1%) and fewer than 2% and 5% of those who remained classified as EL were proficient in the respective subjects. To contextualize these findings, note that 63.7% and 54.1% of fifth graders in Massachusetts who were never classified as EL scored proficient in ELA and mathematics respectively during the same time period.

Research Question 3: One in Five ELs Is Retained in Grade Between Kindergarten and Seventh Grade

To answer Research Question 3—examining the proportion of Massachusetts' students who were retained in grade during the period of analysis—in Table 8, I present longitudinal estimates of the prevalence of in-grade retention in my sample. For those who remained in Massachusetts schools, the rate of in-grade retention is more than double national retention rates. While about 10% of ELs and non-ELs are retained in grade (Kindler, 2002; Planty et al., 2009), 21.6% (n = 1,158) of the 2002 Massachusetts kindergarten cohort was retained at least once during the 8-year period of analysis. The majority of these students (93.8%) were only retained once, and a small proportion of students were retained twice (6%). In Table 9, I show that greater proportions of U.S.-born, low-income, and male students were retained compared to their EL peers who were foreign-born, non-low-income, and female, respectively.

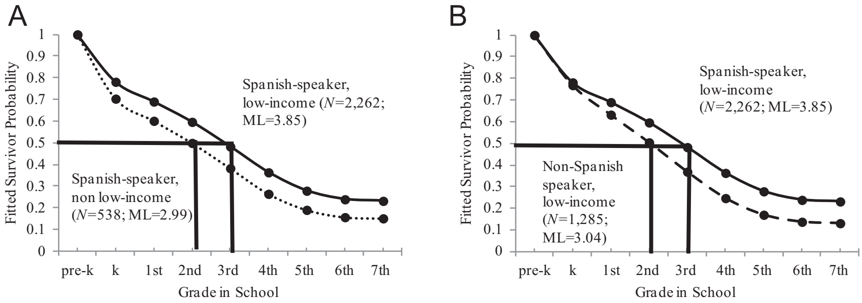

Figure 4. **Estimated survivor functions for prototypical English learners (EL) in the 2002 kindergarten cohort, by grade in school, by income status (Panel A), and by whether a student is a Spanish speaker (Panel B) from fitted Model L.**

Note. Median lifetimes (MLs) overlaid on plot of survivor functions.

In Figure 5, I examined retention cross-sectionally and show that overall retention by grade appears lower than national estimates of retention—with the largest proportion of students being retained in first grade (7.3%) and kindergarten (4.7%). Nationwide, grade retention is linked closely to higher incidences of drop out and poor academic performance (e.g., Bowman-Perrot et al., 2010) and is therefore an important risk factor in the EL population where there is a high prevalence of both (e.g., NCES, 2004). The disparities in the longitudinal and cross-sectional estimates of retention underscore the importance of monitoring EL student performance longitudinally. On the basis of the prevalence of retention in my sample, it is important to obtain separate estimates of time to reclassification and grade of reclassification because the number of years in school will not be a consequence of a linear progression through the grades for retained students.

Discussion

In the present study, I sought to determine whether and when ELs exited language-learning programs and their ability to access the mainstream curricula in U.S. public schools following reclassification. Capitalizing on eight waves of student-level information for a cohort of kindergarten ELs from one Northeast state in the United States, I obtained an accurate picture of students' tenure in language-learning programs and how they fared upon exit—as measured by content-area achievement and retention. Specifically, to estimate the time to and grade of reclassification, I employed a discrete-time survival analysis, which also addressed methodological limitations of most prior research. I estimated time to and grade of reclassification for

Table 8
Patterns of Grade Retention for the 2002–2003 Kindergarten Cohort of Massachusetts ELs (*n* = 5,354)

Retention Pattern	%	*N*
Whole sample		
Retained once or more	21.6	1,158
Never retained	71.3	3,816
Missing retention data	7.1	380
Retained students		
Retained once	93.8	1,086
Retained twice	6.0	69

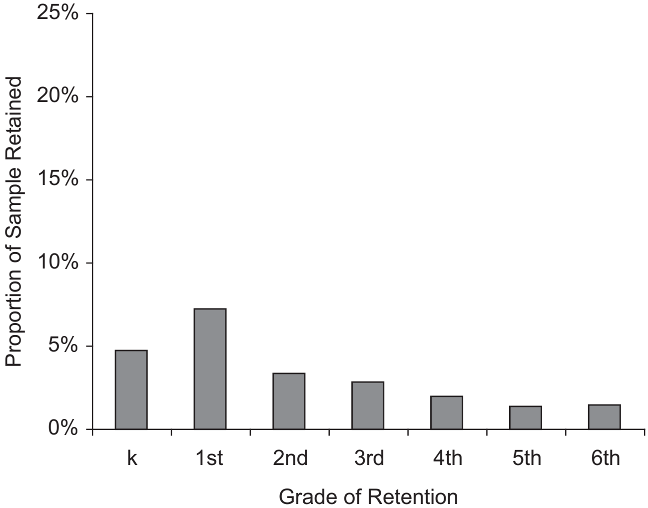

Figure 5. **Proportion of the 2002 kindergarten EL cohort retained each year in grade by grade in school (*n* = 5,354).**

the entire statewide kindergarten cohort as well as for Spanish-speaking, low-income students, the largest and most at-risk EL subgroup.

Four important findings emerge. First, the majority of the 2002 kindergarten EL cohort was reclassified after 3 years in Massachusetts schools (by second grade after accounting for retention). While nearly three quarters of the sample was reclassified at some point during the 8-year period of analysis, by seventh grade, 17% of the sample had not been reclassified. Second, Spanish-speaking, low-income ELs remained classified as EL 1 year longer than their non-Spanish-speaking peers also from low-income backgrounds. Third, large proportions of the sample were retained in grade over the 8-year period (22% compared to 10% nationally), and more than half of reclassified students scored below proficient on statewide content-area assessments. These findings indicate that while ELs may demonstrate grade-level academic proficiency in the early elementary grades, they may still struggle with academic proficiency in the middle and high school grades.

Language and Reclassification Policies Should Reflect Complexities of Language Learning

While the English-only mandates in California, Arizona, and Massachusetts established that ELs should remain in an SEI program for a period of 1 year (Gándara et al., 2010), findings from the present study suggest that reclassification within 1 to 3 years of school entry does not provide an "inoculation" against later academic difficulties.

The finding that large numbers of ELs exited language-learning programs relatively quickly in comparison to other states and districts in the United States with large proportions of ELs—longitudinal estimates of average time to reclassification range from between 5 to 11 years in California (Grissom, 2004; Mitchell et al., 1999; Parrish et al., 2006) and 3 to 6 years in New York City (New York City Board of Education Division of Assessment and Accountability, 2000)—should not be taken as evidence that these students are faring well academically. In fact, their later academic difficulties and a widening performance gap between former ELs and mainstream peers suggests that the skills that ELs need to be successful in early elementary grades are not the same as those required to be successful in the more demanding upper elementary, middle, and high school grades.

Widening performance gaps across the school trajectory have been documented in national data and among the broader group of LM learners. With respect to national data, in Figure 6, I demonstrate a larger gap in performance between reclassified ELs and their monolingual peers in middle grades than elementary grades on the 2009 NAEP (NCES, Institute of Education Sciences, U.S. Department of Education, 2009). Examining the gaps between R-FEP and never-ELs for fourth-grade reading and mathematics compared to the same gap for eighth grade, I show that 29% and 40% of fourth-grade reclassified ELs were proficient in reading and mathematics (versus 34% and 41% of monolingual students), while only 16% and 17% of eighth-grade reclassified ELs were proficient in eighth-grade reading and mathematics (versus 35% and 29% of monolingual students). On the fourth grade MCAS, 56% and 54.6% of R-FEP students were proficient in ELA and mathematics (versus 58.8% and 49.9% statewide "never-ELs"). At seventh grade, the R-FEP/never-EL disparities were larger: 61.7% and 45.9% (R-FEP students) versus 74.2% and 54.9% percent of students never classified

Table 9
Baseline Demographic and School Characteristics for Students Who Were Retained at Least Once Between Kindergarten and Seventh Grade (Versus Not)

	Student Retention Status			
	Retained Once or More (21.6%; n = 1,158)		Never Retained (71.3%; n = 3,816)	
Variable	%	*SD*	%	*SD*
U.S.-born	85.7***	0.35	79.5	0.40
Low-income	78.4***	0.41	69.2	0.46
Immigrant student	14.4	0.35	20.6	0.40
Female	44.9	0.50	49.1**	0.50
Attending urban schools	84.9***	0.36	77.8	0.42
Years as EL	4.9***	2.53	3.7	2.28
Years in school	7.4***	1.29	7.13	1.81

Note. Of the students in the sample, 7.10% are missing retention data (n = 380); therefore, percentages do not add up to full sample (n = 5,354). The t tests compare population means for ELs who were retained at least once in grade, compared to their peers who were never retained. A student has immigrant status when he or she is eligible for the Emergency Immigrant Education Program. To be eligible for this program in Massachusetts, a student must (1) not have been born in any state and (2) not have completed 3 full academic years of school in any state.
$*p < 0.05$. $**p < 0.01$. $***p < 0.001$.

as EL. These results are based on cross-sectional data (i.e., comparisons between different groups of students at one point in time). Future studies should increase the precision of the estimations by investigating how a single group of reclassified students fares academically in later grades, in response to the increasing demands of the middle and high school classroom.

Among LM learners, Kieffer (2008) and Lesaux (2006) found that students' reading skills were on par with native-speaking peers during early elementary years but diverged in upper elementary years. Other studies reported similar patterns among Spanish-speaking ELs and LM students (see Mancilla-Martinez & Lesaux, 2011a; Nakamoto, Lindsey, & Manis, 2007). In effect, word-reading may be the most important predictor of reading achievement in early elementary grades, but beginning in upper elementary grades, vocabulary and comprehension skills play a more central role in reading achievement (e.g., RAND Reading Study Group, 2002; see Kieffer, 2008, for a review). The early-exit students in the present study likely demonstrated on par word-reading skills in the early grades, but their low content-area performance may suggest they struggle with comprehension and vocabulary in the later grades. Students struggling to perform mainstream work in English following reclassification suggests that reclassification in and of itself is not a good predictor of students'

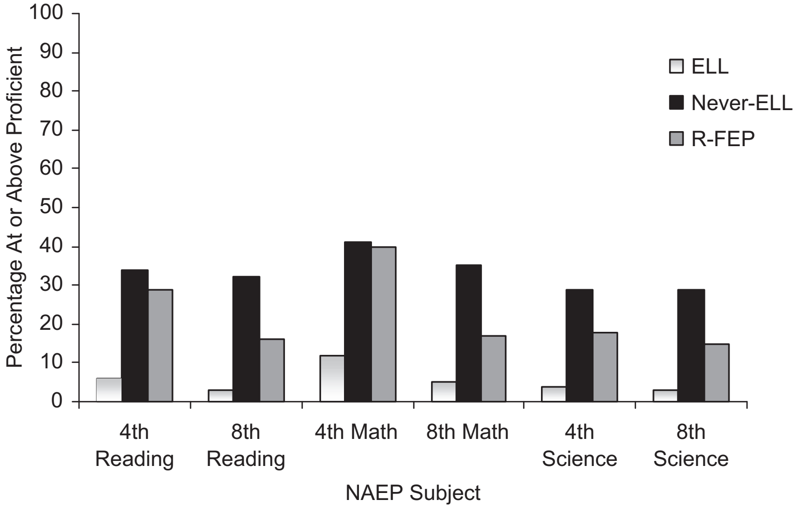

Figure 6. **Percentage of United States public school students scoring at or above proficient on the 2009 NAEP, by language learner status.**

Note. ELL = English language learner; R-FEP = redesignated fluent English proficient; NAEP = National Assessment of Educational Progress.

ability to perform mainstream work in English and a need to re-examine reclassification as an indicator of EL student success.

Prior studies have questioned the use of content-area assessments as a valid and reliable measure of EL content knowledge because the language of the assessment is confounded with the construct the assessment is measuring (for a review, see Abedi, 2002), yet former ELs' lagging performance on the statewide MCAS in the middle grades does not bode well for these learners because in Massachusetts, all students must pass content-area assessments to receive a high school diploma. ELs and their classmates in 26 other states will now be required to pass content assessments to become high school graduates (Center on Education Policy, 2008).

School and Neighborhood Segregation May Explain Lower Reclassification Rates Among Spanish-Speaking ELs

"Triple segregation" of EL students across U.S. schools (Orfield, 2001; Orfield & Lee, 2006; Rios-Aguilar et al., 2012) and in Massachusetts may explain why Spanish-speaking ELs remain in language-learning programs longer than their EL classmates from other language groups. Nationally, attendance in schools *triply segregated* with large proportions of EL, minority, and low-income students is associated with lower achievement for all students in those schools (Orfield, 2001; Orfield & Lee, 2006; Rios-Aguilar & Gándara, 2012) and particularly for ELs (Ruiz-de-Vasco, Fix, & Chu Clewell, 2000) and Latino students for whom school segregation has increased at the fastest pace (Orfield, 2001). In Massachusetts, the average EL attends a triply segregated school (see Figure 1), but Spanish-speakers are more likely to be clustered in high-poverty schools and districts compared to their EL peers who speak Asian languages at home (Owens, 2010).

Less exposure to native English speakers may be one mechanism through which school and neighborhood segregation may be linked to lower reclassification rates for Spanish-speaking ELs. Prior empirical studies have established that coming from a low-income home and status as a second language learner put students at greater risk for reading difficulties largely because of the deleterious effects of poverty and the amount of English language exposure in the home and community (see Kieffer, 2011; Mancilla-Martinez & Lesaux, 2011b, for reviews). Linguistic isolation has been linked to ELs' low achievement in Arizona (Lillie et al., 2010; Rios-Aguilar et al., 2012) and California (Gándara et al., 2003), whose EL population is mostly Spanish-speaking (84% and 83%, respectively). In addition, Gándara et al. (2003) cite lack of peer English language models and exposure to higher achieving classmates, inadequate resources, and lack of highly qualified classroom teachers as the largest threats of segregation to ELs' language development. In Massachusetts, the deleterious effects of segregation on ELs' academic outcomes are likely to be more pronounced for Spanish-speaking ELs as they comprise the largest language group in the state and tend to be clustered in high-poverty urban schools and districts. In Boston, the largest EL district in the state, 61% of EL students speak Spanish as a home language and one in five students across the district is a Spanish-speaker (Tung et al., 2009). Tung et al. (2009) found that Spanish-speaking ELs were at greater risk of educational failure as measured by higher dropout rates, lower engagement, and lower MCAS scores compared to their EL peers from other language groups.

Longitudinal Monitoring System Would Improve Current EL Accountability System

The present study illustrates how cross-sectional data can mask important EL performance trends and highlights the need for state and federal investment in longitudinal monitoring systems. First, consider the largest group of students in my study, those reclassified by the end of third grade; under the current state and federal monitoring systems, which follows EL performance for 2 years after exit, their academic performance would be monitored until the end of their fifth grade year. That means that by the time these students enter the middle school grades and begin to encounter more demanding academic content, they have now lost both the EL and

R-FEP labels and respective learning supports (i.e., specialized language-learning services for EL students and academic progress monitoring for R-FEP students). They are now expected to perform on par with mainstream grade-level peers absent additional support, but findings from this study suggest that many of these students are ill prepared to do so. Second, annual (i.e., cross-sectional) retention rates for ELs were within national averages (e.g., Figure 5), but when examined over time, the proportion of ELs who experienced retention over their schooling trajectory was double the national norm. Third, nearly one in five ELs in the sample was never reclassified during the 8-year period of analysis, but the state only collects data on the number of years that ELs remain in Massachusetts schools, not the number of years as EL. Evidence of a widening performance gap among reclassified ELs and their mainstream peers from large, representative national and statewide samples of ELs suggests the need for intervention at the federal and state level.

Implications, Limitations, and Directions for Future Research

An important implication of the present study is that the majority of students in the kindergarten EL cohort studied remained in the state during their schooling trajectories, was reclassified within 3 years of school, and experienced later academic difficulties in both English and mathematics. In effect, standards, assessments, language-learning support, and monitoring systems at the state and district level must reflect the reality that academic challenges facing younger ELs are different from those that older learners will encounter.

There are four limitations to these findings. First, the data are observational in nature and cannot support causal inferences about the impact of reclassification of ELs into mainstream classrooms on subsequent academic performance. Second, Massachusetts has proportionally fewer ELs compared to the other states with English-only mandates, which may limit the external validity of my findings. Third, controlling for income when examining time to reclassification for Spanish-speakers does not control for parental education or social capital, which may be more predictive of immigrant student achievement than income alone (see Kao & Taggart Rutherford, 2007, for a review). Last, the EL population is heterogeneous (see Goldenberg, Rueda, & August, 2006, for a review) and a blanket policy regarding the time or grade by which students should be exited from language-learning programs is inappropriate.

Three directions for stakeholders concerned with ELs' academic outcomes emerge. First, researchers should investigate the relationship between long-term EL status and levels of academic English proficiency, EL program quality, lack of exposure to mainstream curricula and classmates, and NCLB accountability incentives. Researchers should examine factors that promote or hinder the achievement of at-risk EL subgroups such as Spanish-speaking, low-income students. Second, this study adds further support to the call for the replacement of the current 2-year monitoring system and unstable EL subgroup definition with an accountability system that monitors EL performance across students' entire school trajectory with a stable "EL at school entry" designation. Accordingly, resources must be allocated for states to revamp their data management systems to monitor EL performance over time. Currently, few states have the technical capacities to monitor student performance longitudinally. Third, districts and schools must have funds to intervene when students remain in EL programs beyond 6 years or when reclassified ELs experience language-based academic difficulties. In sum, language-learning services must be reconceptualized—not as an early inoculation against later academic difficulties—but as an ongoing system of support and progress monitoring that acknowledges the changing demands of the middle and high school classroom.

Note

The author wishes to express appreciation to Catherine Snow, John Willett, and Nonie Lesaux for their contributions to the study design and analysis. The author expresses gratitude to Carrie Conaway, Kendra Winner, Lien Hoang, and Robert Lee at the Massachusetts Department of Elementary and Secondary Education (MDESE) and Nealia Khan for their assistance with data access and construction of the longitudinal dataset. The author is grateful to leaders of EL departments across the state for their insight into reclassification practices in their respective districts. This work was funded with generous support from the American Institutes for Research, the Gordon M. Ambach Fellowship, and the State Farm Companies Foundation Fellowship.

References

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometric issues. *Educational Assessment*, *8*(3), 231–257.

Arias, M. B., & Faltis, C. (2012). *Implementing educational language policy in Arizona: Legal, historical and current practices in SEI*. Bristol, UK: Multilingual Matters.

Batalova, J., Fix, M., & Murray, J. (2007). *Measures of change: The demography and literacy of adolescent English language learners—A report to the Carnegie Corporation of New York*. Washington, DC: National Center on Immigration Integration Policy, Migration, Migration Policy Institute.

Bowman-Perrot, L. J, Herrera, S., & Murry, K. (2010). Reading difficulties and grade retention: What's the connection for English language learners? *Reading & Writing Quarterly*, *26*, 91–107.

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

Castañeda v. Pickard, 648 F.2d 989 (U.S. Ct. App. 1981).

Center on Education Policy. (2008). *State high school exit exams: Moving toward end-of-course exams*. Washington, DC: Author.

Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, *21*(4), 617–641.

Cosentino de Cohen, C., Deterding, N., & Chu Clewell, B. (2005). *Who's left behind? Immigrant children in high and low LEP schools* (policy report). Washington, DC: Urban Institute.

de Jong, E. (2004). After exit: Academic achievement patterns of former English language learners. *Educational Policy Analysis Archives*, *12*(50), 1–20.

Flores, E., Painter, G., Harlow-Nash, Z., & Pachon, H. (2009). *Qué pasa? Are English language-learning students remaining in English learning classes too long?* (policy brief). Los Angeles: Tomas Rivera Policy Institute.

Fry, R., & Gonzales, F. (2008). *One-in-five and growing fast: A profile of Hispanic public school students*. Washington, DC: Pew Hispanic Center.

Gándara, P., Losen, D., August, D., Uriarte, M., Gómez, M. C., & Hopkins, M. (2010). Forbidden language: A brief history of U.S. language policy. In P. Gándara & M. Hopkins (Eds.), *English learners and restrictive language policies* (pp. 20–49). New York: Teachers College Press.

Gándara, P., & Merino, B. (1993). Measuring the outcomes of LEP programs: Test scores, exit rates, and other mythological data. *Educational Evaluation and Policy Analysis*, *15*(3), 320–338.

Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, *11*(36). Retrieved from http://epaa.asu.edu/epaa/v11n36/

Goldenberg, C., Rueda, R. S., & August, D. (2006). Sociocultural influences on the literacy attainment of language-minority children and youth. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 365–414). Mahwah, NJ: Erlbaum.

Grissom, J. B. (2004). Reclassification of English language learners. *Education Policy Analysis Archives*, *12*(26).

Hakuta, K., Goto Butler, Y., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* (Policy report). Berkeley: University of California Linguistic Minority Research Institute.

Kao, G., & Taggart Rutherford, L. (2007). Does social capital still matter? Immigrant minority disadvantage in school-specific social capital and its effects on academic achievement. *Sociological Perspectives*, *50*(1), 27–52.

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*(4), 851–868.

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.

Kieffer, M. J., Lesaux, N. K., & Snow, C. E. (2008). Promises and pitfalls: Implications of NCLB for identifying, assessing, and educating English language learners. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform* (pp. 57–74). Thousand Oaks, CA: Kindler.

Kindler, A. L. (2002, October). *Survey of the states' limited English proficient students and available educational programs and services 2000-2001 summary report*. Silver Spring, MD: National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs. Retrieved from http://www.ncela.gwu.edu/policy/states/reports/seareports/0001/sea0001.pdf

Lau v. Nichols, 414 U.S. 56 (1974). Retrieved February 13, 2011, from http://www2.Ed.gov/about/offices/list/ocr/EL/lau.Html

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

Lillie, K. E., Markos, A., Estrella, A., Nguyen, T., Trifiro, A., Arias, M. B.,… Pérez, K. (2010). *Policy in practice: The implementation of Structured English Immersion in Arizona*. Los Angeles: The Civil Rights Project. Retrieved from http://civilrightsproject.ucla.edu/research/k-12-education/language-minority-students/policy-in-practice-the-implementation-of-structured-english-immersion-in-arizona/lillie-policy-practice-sei-2010.pdf

Linquanti, R. (2001). *The redesignation dilemma: Challenges and choices in fostering meaningful accountability for English learners* (No. Policy Report 2000-1). Santa Barbara: University of California Linguistic Minority Research Institute.

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

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.

Massachusetts Department of Elementary and Secondary Education. (2003). *SIMS Version 2.0 Data standards handbook for the Massachusetts Student Information Management System*. Malden: Author.

Massachusetts Department of Elementary and Secondary Education. (2009). *Guidance on using MEPA results to plan sheltered English immersion (SEI) instruction and to make reclassification decisions for limited English proficient (LEP) students*. Retrieved from www.doe.mass.edu/mcas/mepa/2009/guidance.doc

Massachusetts Department of Elementary and Secondary Education. (2012). *Transitioning English language learners in Massachusetts: An exploratory data review*. Malden: Author.

Mitchell, D. E., Destino, T., Karam, R. T., & Colón-Muñiz, A. (1999). The politics of bilingual education. *Educational Policy*, *13*(1), 86–103.

Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading & Writing: An Interdisciplinary Journal*, *20*, 691–719.

National Center for Education Statistics. (2004). *Language minority learners and their labor market indicators: Recent trends*. Washington, DC: U.S. Department of Education.

National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. (2009). Washington, DC. Retrieved from http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx

New York City Board of Education Division of Assessment and Accountability. (2000). *English language learner (EL) subcommittee research studies progress report*. Retrieved from http://schools.nyc.gov/daa/reports/EL_Research_Studies.pdf

Orfield, G. (2001). *Schools more separate: Consequences of a decade of resegregation*. Cambridge, MA: Harvard Civil Rights Project. Retrieved from http://www.civilrightsproject.ucla.edu/research/deseg/separate_schools01.php

Orfield, G., & Lee, C. (2006). *Racial transformation and the changing nature of segregation*. Cambridge, MA: The Civil Rights Project, Harvard University.

Owens, A. (2010). *In the aftermath of Question 2: Students with limited English proficiency in Massachusetts*. Boston: Massachusetts Office for Refugees and Immigrants. Retrieved from http://www.hks.harvard.edu/centers/rappaport/stu dent-opportunities/public-policy-summer-fELowship/reports-and-publications/in-the-aftermath-of-question-2-students-with-limited-english-proficiency-in-massachusetts

Parrish, T. B., Merickel, A., Perez, M., Linquanti, R., Socias, M., Spain, A., et al. (2006). *Effects of the implementation of Proposition 227 on the education of English learners, K-12 findings from a five year evaluation* (No. AB 56 and AB 1116). Palo Alto, CA: American Institutes for Research.

Planty, M., Hussar, W., Snyder, T., Kena, G., Kewal Ramani, A., Kemp, J., Bianco, K., & Dinkes, R. (2009). *The Condition of Education 2009* (NCES 2009-081). Washington, DC: National Center for Education Statistics.

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).

Ramirez, D. J., Yuen, S. D., & Ramey, D. R. (1991). *Final report: Longitudinal study of structured immersion strategy, early-exit, and late-exit transitional bilingual education programs for language-minority children*. San Mateo, CA: Aguirre International.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Corporation.

Reynolds, A. J., Chen, C.- C., & Hebers, J. E. (2009). *School mobility and educational success: A research synthesis and evidence on prevention*. Paper commissioned by the National Academy of Sciences Committee on the Impact and Change in the Lives of Young Children, Schools, and Neighborhoods, Washington, DC, June 29–30.

Rios-Aguilar, C., & Gándara, P. (2012). *Horne v. Flores* and the future of language policies. *Teachers College Record*, *114*, 1–13.

Rios-Aguilar, C., González-Canché, M. S., & Sabetghadam, S. (2012). Evaluating the impact of restrictive language policies: The Arizona 4-hour English language development block. *Language Policy*, *11*, 47–80.

Rolstad, K., Mahoney, K. S., & Glass, G. V. (2005): Weighing the evidence: A meta-analysis of bilingual education in Arizona. *Bilingual Research Journal: The Journal of the National Association for Bilingual Education*, *29*(1), 43–67.

Ruiz-de-Vasco, J., Fix, M., & Chu Clewell, B. (2000). *Overlooked and underserved: Immigrant students in U.S. secondary schools*. Washington, DC: The Urban Institute.

Rumberger, R. W., & Tran, L. (2010). State language policies, school language practices and the English learner achievement gap. In P. Gándara & M. Hopkins (Eds.), *English learners and restrictive language policies* (pp. 86–101). New York: Teachers College Press.

Scarcella, R. (2003). *Academic English: A conceptual framework* (University of California Linguistic Minority Research Institute Technical Report 2003–1). Retrieved from http://lmri.ucsb.edu/publications/03_scarcELa.pdf

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

Slama, R. B. (2011). A longitudinal analysis of academic English proficiency outcomes for adolescent English language learners in the United States. *Journal of Educational Psychology*, *104*(2), 265–285.

Tung, R., Uriarte, M., Diez, V., Lavan, N., Augusti, N., Karp, F., et al. (2009). *English learners in Boston public schools: Enrollment, engagement and academic outcomes, AY2003-AY2006*. Boston: Mauricio Gastón Institute for Latino Community Development and Public Policy.

U.S. Department of Education. (2009). *Fact sheet: NCLB provisions ensure flexibility and accountability for limited English proficient students*. Retrieved December 13, 2009, from http://www.ed.gov/nclb/accountability/schools/factsheet-english.html

U.S. Department of Education. (2012). *Arizona Department of Education: Resolution agreement*. Retrieved from http://www.azed.gov/public-relations/files/2012/08/executed-resolutionqqagreement-08-31-12.pdf

Williams, M. L. (1991). *Policy update on schools' obligations towards national origin minority students with limited-English proficiency*. Washington, DC: U.S. Department of Education.

The Working Group on ELL Policy. (2011). *Improving educational outcomes for English language learners: Recommendations for ESEA reauthorization*. Retrieved from http://ELpolicy.org/wp-content/uploads/PolicyBrief.pdf

Manuscript received April 19, 2012
Final revision received May 28, 2013
Accepted September 30, 2013