

To Be or Not to Be EL: An Examination of the Impact of Classifying Students as English Learners

Hana M. Umansky
University of Oregon

Across the United States, students who are deemed not to be proficient in English are classified as English learners (ELs). This classification entitles students to specialized services but may also result in stigmatization and barriers to educational opportunity. This article uses a regression discontinuity design to estimate the effect of EL classification in kindergarten on students' academic trajectories. Furthermore, it explores whether the effect of EL classification differs for students in English immersion versus bilingual programs. I find that among language-minority students who enter kindergarten with relatively advanced English proficiency, EL classification results in a substantial negative net impact on math and English language arts test scores in Grades 2 through 10. This effect, however, is concentrated in English immersion classrooms.

Keywords: *immigration/immigrants, longitudinal studies, educational policy, regression discontinuity, bilingual/bicultural*

IN SCHOOLS across the United States, students who speak a language other than English at home have their English assessed when they first enter school. Those who score below a set threshold are classified as “English learners” (ELs). Those who score at or above the threshold are considered initially fluent English proficient (IFEP).¹ IFEP students are mainstream students, receiving the same services as students who speak only English. Although IFEP classification is neutral, EL classification carries with it important legal implications for the provision of services and treatments. EL classification is tightly linked to the provision of English language development (ELD) instruction, as well as other programmatic effects such as “sheltered” academic content instruction and, in some locales, instruction in students’ home languages. In addition to these programmatic treatments generated by EL classification, research suggests that EL classification may also generate status effects such as diminished teacher expectations or social stigma.

With seemingly positive treatments and seemingly negative treatments, it is unclear how EL

classification impacts students and under what circumstances. This article investigates these questions. It examines the impact of EL classification in kindergarten on a set of students’ medium- to long-term educational outcomes using data from a large urban school district in California.

The article takes advantage of a natural experiment that occurs at the cusp of EL–IFEP classification. As I will show, kindergarten students who score at or just above the EL threshold on the English assessment are indistinguishable from those who score just below the threshold in every way except their subsequent language classification. Any average differences in the educational trajectories of these two groups are the result of their language classification status and associated treatments.

I find that students at the margin of EL–IFEP classification in kindergarten (I refer to these students as marginal or cusp students) and who are classified as EL, as compared with IFEP, have significantly lower test scores in math and English language arts (ELA) in Grades 2 through 10. Point estimates suggest that this gap is

sizable in second grade and grows slowly in magnitude as students progress through elementary and secondary school. By the secondary level, the effect is equivalent to roughly one quarter of California's achievement gap between Latinos and Whites in ELA and math. These results are only generalizable to students who enter kindergarten with relatively advanced English proficiency and may also differ in different school districts and time periods. However, I argue that students at the threshold are an important group of students from a policy perspective. This is because they are akin to the "canary in the coalmine" in that they have comparatively little to gain from EL programmatic services rendering them particularly vulnerable to negative treatments associated with EL classification.

Although EL classification has an average negative effect on students at the threshold, this district offers a unique opportunity to examine the impact of such classification in each of the four linguistic instructional programs—one traditional English immersion (EI) program and three two-language programs. Due to differences in program goals and design, both programmatic and status treatments of EL classification may differ in EI versus two-language classrooms. I find that among students at the cusp of EL–IFEP classification, EL classification negatively impacts students in all English instructional environments but is neutral for students in at least two of the three two-language programs. I conclude that EL classification is consequential for marginal students but that the effects of EL classification may be malleable to school and district practice and policy.

EL Classification and Treatments

Title I and Title III of the recently reauthorized Elementary and Secondary Education Act stipulate that states that receive federal education funds must identify who among their students are ELs and annually assess how ELs, as a subgroup, do on English language proficiency assessments and academic content assessments. EL classification likely triggers two types of treatments: first, programs and services designed to meet ELs' unique educational needs; second, changes in their social status relative to non-ELs. I refer to the former as programmatic treatments and the latter as status treatments.

EL Programmatic Treatments

On the programmatic side, federal education code stipulates that EL-classified students must be provided ELD instruction as well as meaningful access to grade-level academic content (Ramsey & O'Day, 2010). ELD is direct instruction in the English language, designed to advance ELs' English competency and facilitate successful participation in academic subject areas in school (Saunders, Goldenberg, & Marcellotti, 2013). Academic content supports for ELs typically consist of (a) content area instruction in English using techniques to increase accessibility (often called sheltered instruction) and/or (b) content instruction in students' home language (here called two-language instruction to differentiate it from specific programs entitled bilingual). Sheltered instruction is grade-level academic content instruction that employs modifications for ELs such as integrating language objectives into class, using visual aids, and providing extra time for practice (Goldenberg & Coleman, 2010).

Two-language instruction is academic content instruction that is delivered in part, or in whole, in students' home language. There are three main two-language instructional models currently in practice, all of which are offered in the school district examined here. Transitional and maintenance bilingual instructional models are designed specifically for ELs. Transitional bilingual (TB) programs are typically 3 to 4 years in duration and focus on using the home language to support English acquisition and access to curricular content. Maintenance bilingual (MB) programs are longer in duration and prioritize full bilingualism in English and the home language. The third two-language model is dual immersion (DI) and it differs from the prior two primarily in terms of student composition. DI programs include both language-minority students and English-only speakers (EOs) with the goal that both groups develop proficiency in both languages (Billings, Martin-Beltrán, & Hernandez, 2010).

EL Status Treatments

EL classification is not designed to impact individuals' social status, but there is wide acknowledgment that it often does. Both the classification itself and the services that accompany the classification are often stigmatized (Dabach,

2014; Valdés, 1998b; Valenzuela, 1999). Prior research suggests that teachers and peers may associate ELs with negative stereotypes including being less academically able, passive, unmotivated, and less socially integrated (Gougeon, 1993; Spack, 1997; Vollmer, 2000). Other research suggests that some EL-classified students internalize stigma and develop feelings of academic and social inferiority (Dabach, 2014; Thompson, 2015; Valenzuela, 1999).

The Impact of Labels

Theory on the impact of labels suggests that labels bring with them a set of treatments, including both intentional treatments—such as services—and unintentional treatments—such as altered perceptions of the individual based on the label. Each of these treatments may have an impact on individuals' outcomes (Goffman, 1963; Link, Cullen, Struening, Shrout, & Dohrenwend, 1989; Link & Phelan, 2013; Scheff, 1970). Studies suggest that intended treatments often have positive effects whereas unintended treatments often have negative effects (Link et al., 1989).

In education, labels that identify students' real or perceived ability or achievement have been found to have a meaningful impact on student outcomes (Rist, 1977; Rosenthal & Jacobson, 1968). In two recent studies, labeling students based on test scores is found to impact both student achievement and college going (Domina, Penner, & Penner, 2014; Papay, Murnane, & Willett, 2010).

Special education labels share several key attributes with the EL label and have been the subject of study over the past several decades. Like EL classification, students classified with special education labels are identified by their deviation from a social norm (Becker, 1963; Link & Phelan, 2013) in terms of English proficiency in the case of the EL label (Pennycook, 2002) and in terms of mental, physical, social, and/or emotional development in the case of the special education label (McDermott, 1993). In addition, both EL and special education classifications give rights to students and responsibilities to education systems to provide specialized services to meet students' learning needs. Research on the impact of special education labels has

shown that they negatively impact teachers' perceptions and expectations of students (Bianco, 2005) and that they negatively alter peers' perceptions and treatment of students (Bak, Cooper, Dobroth, & Siperstein, 1987). Furthermore, evidence suggests that special education–classified students experience stigmatization (Higgins, Raskind, Goldberg, & Herman, 2002; Jones, 1971) and negative academic outcomes (Morgan, Frisco, Farkas, & Hibel, 2008; Sullivan & Field, 2013).

Hypothesizing the Impact of EL Classification

There are compelling arguments as to why classifying students as ELs might be beneficial for students and there are equally compelling arguments as to why it might be harmful. In all likelihood, both sets of arguments operate in tandem, with some aspects of EL classification helping students and others causing harm (Link et al., 1989; Link & Phelan, 2013). Whether the net impact of EL status is positive, negative, or neutral depends on the relative strength of the positive and negative effects. This net impact is unlikely to be universal but rather is affected by multiple factors related to individual student characteristics as well as characteristics of the school, community, and treatments and services given to ELs.

Why EL Classification May Improve Student Outcomes

Although research on the effectiveness of EL services is limited (Goldenberg, 2013), there is evidence that many of the programmatic services triggered by EL classification are beneficial. Meta-analyses of research on bilingual education suggest that instruction in a student's home language benefits the development of English proficiency and, at a minimum, does not result in inferior academic outcomes (August & Shanahan, 2006; Slavin, Madden, Calderón, Chamberlain, & Hennessy, 2011; Thomas & Collier, 2002). Other EL services that have been evaluated and found to be beneficial include adapting content instruction to focus on academic vocabulary, integrating English literacy and oracy instruction into content area teaching, offering a dedicated

block of leveled ELD (Saunders, Foorman, & Carlson, 2006; Saunders et al., 2013), and scaffolding instruction for ELs (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Baker et al., 2014; Kim et al., 2011; Walqui, 2006). Teacher training and professional development focused on EL instruction has also been found to benefit ELs (Master, Loeb, Whitney, & Wyckoff, 2012). Finally, specialized services and classes for ELs can create supportive learning environments that aid student learning and growth (Chang et al., 2007; Harklau, 1994; Valentino & Reardon, 2015).

Why EL Classification May Hurt Student Outcomes

Although many of the targeted programs and services for ELs have themselves been linked to improvements in EL learning, there is a growing body of evidence that the EL classification system or aspects of it can stymie learning.

Widespread associations of EL classification with ideas of inferiority, inability, and remediation can result in EL students internalizing negative self-concepts (Dabach, 2014; Thompson, 2015), which, in turn, may negatively impact student learning (Link & Phelan, 2013; Steele, 1997). Teachers may adapt their behavior based on conceptions of EL inferiority by diminishing the rigor of curricular content (Dabach, 2014) or school administrators may respond by placing EL students in lower track classes (Estrada, 2014; Kanno & Kangas, 2014).

On the programmatic side, services that are designed to help students learning English may carry with them unintended consequences that penalize students academically. First, the provision of specialized EL services can result in isolation from English-speaking peers in separate schools or classrooms with little opportunity to speak or hear English during the day aside from interactions with teachers (Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Gifford & Valdés, 2006; Katz, 1999). Second, the provision of EL services may crowd out participation in mainstream academic classes. Documentation of EL course-taking suggests that classes for ELs such as ELD may substitute rather than complement core academic classes (Umansky, 2015). Closely related to this, EL classification may be linked to tracking practices that

limit access to full academic participation or advanced academic courses (Estrada, 2014; Kanno & Kangas, 2014).

Linguistic Instructional Program as a Moderator of the Impact of EL Classification

With multiple treatments, some potentially positive and others potentially negative, the impact of EL classification on students' outcomes is unlikely to be uniform; instead, it is likely to vary based on student, school, and social contextual factors (Callahan, Wilkinson, & Muller, 2008, 2010; Robinson-Cimpian & Thompson, 2014).

The most logical factor that may influence the net impact of EL classification is the services that students receive as ELs. A school that offers high quality and well-targeted services to ELs is likely to have stronger positive effects of EL classification than a school with low quality and inappropriately targeted services. One important way in which service provision for ELs varies is linguistic model. In the district studied here, ELs can enroll in one of four instructional programs: EI, TB, MB, or DI.

In two-language classrooms, EL classification, compared with IFEP classification, may be more beneficial/less detrimental than it is in English-only classrooms. First, EL status may be less stigmatized in two-language classrooms, which typically value bilingualism, and home languages and cultures (Crawford, 1989). Second, as the population of focus in two-language classrooms, EL-classified students may be less vulnerable to tracking or crowding-out of academic content, compared with EL students in EI classrooms (Callahan et al., 2008). This second point may apply more specifically to TB and MB programs and less to DI programs because the latter serves both ELs and a frequently more vocal and politically powerful English-speaking population (Valdés, 1998a). These characteristics of two-language classrooms could result in different net impacts of EL classification across programs.

Prior Research on the Impact of EL Classification

Existing studies begin to shed light on the impact of EL classification on student outcomes. Many regression-based studies include EL as a standard

control variable. Although these studies typically find a significant negative point estimate on the EL variable, these estimates should not be considered causal estimates of EL classification. This is because ELs differ in many unobservable ways from non-ELs, generating selection bias. Therefore, the EL point estimates in standard regression analyses are unlikely to isolate any EL classification effect.

A smaller group of studies use quasi-experimental methods to more rigorously identify the effect of being an EL. A set of studies using propensity score matching suggests that the impact of EL status is variable, as posited above (Callahan et al., 2008, 2010; Callahan, Wilkinson, Muller, & Frisco, 2009). These studies compare students receiving EL services with similar students not receiving services and find that EL services are associated with inferior educational outcomes among students with higher English language proficiency levels, students in schools with fewer ELs, and students who have been in the United States longer and are from more socioeconomically advantaged backgrounds. By contrast, EL services are associated with superior educational outcomes among students with the opposite characteristics.

A second set of studies uses regression discontinuity (RD) to examine the impact of being reclassified out of EL status (Carlson & Knowles, 2016; Robinson, 2011; Robinson-Cimpian & Thompson, 2014). Two of these studies find that there is often no discernible impact of EL status compared with reclassified status for students at the cusp of reclassification (Robinson, 2011; Robinson-Cimpian & Thompson, 2014). The third finds a negative effect of remaining classified as an EL on standardized test scores, graduation, and postsecondary enrollment among students at the cusp or reclassification (Carlson & Knowles, 2016). The present study is the first to examine the causal impact of ever having been classified as an EL versus never having been classified as an EL, as such identifying the impact of EL classification on cusp students.

Data and Method

Data

This article uses longitudinal data from a large, urban school district in California. In the district, students whose families indicate that

they speak a language other than English at home must take the California English Language Development Test (CELDT) upon entry into the district. The test is comprised of four subtests—reading, writing, speaking, and listening. Kindergarten students must meet minimum scores on the listening and speaking subtests and the combined score (the cut-scores for each fluctuate somewhat by year) to be classified as IFEP and placed into mainstream services. Students who do not meet all of these benchmarks are classified as ELs and receive EL services and associated treatments.

The sample includes EL and IFEP students from nine kindergarten cohorts who entered the district in fall 2002 through fall 2010; covariates and outcomes are measured from 2002 through 2012. Table A in the appendix (available in the online version of the journal) shows the number of students in each cohort as they move through academic grades. The primary outcomes of interest are student scores on the California Standards Tests (CSTs). Until 2014, when a new Common Core State Standards-aligned test was implemented, every student in California in Grades 2 to 11 took a math and an ELA CST every spring. CSTs are offered exclusively in English. For ELA, I have sufficient years of data to measure outcomes from second through tenth grade. For math, I measure outcomes from second through seventh grade. I omit scores beyond seventh grade because, beginning in the eighth grade, students take differentiated math tests and I do not want to confound test score with test type.

I limit the sample to those students for whom I have CELDT scores in kindergarten (87% of students), as this is the test that determines student language status. I did not impute missing data. In total, the analytic sample consists of 18,208 students and 106,497 student-year observations. Table 1 presents descriptive statistics on the total analytic sample and by initial language classification.

EL services in the district are similar to those in many parts of the country. They consist of 30 minutes of daily ELD instruction as well as sheltered academic instruction in which teachers adapt methods and content for EL accessibility. In addition, the district offers optional two-language instruction.

TABLE 1

Descriptive Statistics of Analytic Sample, for Total Sample and by Kindergarten Language Classification

		Total	EL	IFEP	<i>F</i> -test comparing means
Student characteristics					
Female	Percent	49.5%	48.6%	56.1%	***
	Count	9,019	7,745	1,274	
Born in the United States	Percent	78.1%	77.3%	83.4%	***
	Count	14,218	12,323	1,895	
Ever special education identified	Percent	12.2%	13.3%	4.2%	***
	Count	2,218	2,122	96	
English home language	Percent	12.1%	9.7%	28.8%	***
	Count	2,202	1,548	654	
Spanish home language	Percent	34.4%	35.9%	23.7%	***
	Count	6,261	5,723	538	
Cantonese/Mandarin home language	Percent	35.8%	37.6%	23.6%	***
	Count	6,526	5,990	536	
Initial CELDT scores					
Overall	Average	-0.98	-1.17	0.34	***
Listening/speaking	Average	-0.35	-0.57	0.81	***
Listening	Average	-0.64	-0.81	0.79	***
Speaking	Average	-0.41	-0.55	0.87	***
Reading	Average	-0.35	-0.47	0.70	***
Writing	Average	-0.58	-0.66	0.14	***
Second grade CST scores					
Math	Average	0.24	0.17	0.79	***
ELA	Average	0.08	-0.01	0.71	***
Initial instructional program					
Enrolled in English immersion in K	Percent	56.7%	55.1%	67.5%	***
	Count	10,319	8,786	1,533	
Enrolled in dual immersion in K	Percent	11.1%	10.0%	18.4%	***
	Count	2,018	1,601	417	
Enrolled in transitional bilingual in K	Percent	15.9%	17.4%	5.9%	***
	Count	2,902	2,768	134	
Enrolled in maintenance bilingual in K	Percent	12.5%	13.5%	5.9%	***
	Count	2,285	2,150	135	
<i>n</i>		18,208	15,936	2,272	

Note. The *F*-test column gives the significance level on the *F* statistic, calculated using one-way ANOVA. CELDT scores are standardized by the total-sample standard deviation and centered at the respective cut-scores for test and year. CST scores are standardized and centered using state means and standard deviations by test, year, and grade. EL = English learner; IFEP = initially fluent English proficient; CELDT = California English Language Development Test; CST = California Standards Test; ELA = English language arts; K = kindergarten.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Parents choose between the district's four linguistic instructional programs. The largest program, serving 57% of the sample, is an EI program in which ELs are placed in general education classrooms with monolingual English speakers. The district also has three two-language programs including (a) TB, kindergarten-Grade 3

(K-3), serving 16% of the sample; (b) MB, K-5 or above, serving 13% of the sample; and (c) DI, also K-5 or above, which serves 11% of the sample. All schools offer EI and many, particularly at the elementary level, offer one, and in rare instances two, of the two-language programs. Parents list their preferred schools and programs

TABLE 2

Characteristics of Analytic Sample, by Fluency Status and Initial Instructional Program Enrollment

Student characteristics	English immersion		Dual immersion		Transitional bilingual		Maintenance bilingual	
	EL	IFEP	EL	IFEP	EL	IFEP	EL	IFEP
Latino	25.0%	20.4%	72.9%	38.8%	42.9%	44.0%	49.3%	43.6%
Chinese	43.2%	33.4%	13.5%	19.9%	51.8%	48.0%	44.2%	32.9%
Other ethnic group	31.8%	46.1%	13.5%	41.2%	5.3%	8.0%	6.6%	23.6%
Female	47.7%	56.0%	51.8%	55.4%	50.3%	59.3%	50.2%	55.0%
Born in the United States	79.6%	82.5%	86.9%	90.2%	75.5%	79.3%	73.6%	76.4%
Ever special education identified	11.9%	4.4%	14.8%	3.6%	11.4%	2.7%	13.8%	6.4%
English home language	13.2%	29.4%	10.1%	36.2%	3.3%	11.3%	3.0%	20.0%
Spanish home language	22.8%	16.8%	72.1%	37.4%	42.9%	41.3%	50.0%	40.7%
Cantonese/Mandarin home language	37.8%	24.3%	10.6%	12.9%	50.0%	40.0%	43.0%	29.3%
Second grade CST score—ELA	0.12	0.81	-0.46	0.47	0.09	0.59	-0.19	0.51
Second grade CST score—math	0.24	0.85	-0.24	0.66	0.33	0.72	0.04	0.57
Initial CELDT overall score	-0.98	0.35	-1.14	0.41	-1.37	0.21	-1.38	0.23
Initial CELDT listening/speaking score	-0.40	0.80	-0.57	0.90	-0.74	0.73	-0.69	0.76
Initial CELDT listening score	-0.64	0.83	-0.73	0.75	-1.00	0.60	-1.03	0.66
Initial CELDT speaking score	-0.34	0.88	-0.49	0.97	-0.78	0.67	-0.83	0.64
Initial CELDT reading score	-0.27	0.80	-0.66	0.47	-0.61	0.34	-0.71	0.65
Initial CELDT writing score	-0.51	0.20	-0.76	-0.04	-0.71	0.19	-0.78	0.27
Total years in dual immersion	0.04	0.01	3.72	3.71	0.10	0.03	0.11	0.03
Total years in English immersion	5.58	5.98	1.34	1.40	2.74	4.25	2.39	3.79
Total years in transitional bilingual	0.02	0.00	0.01	0.00	2.90	2.49	0.11	0.10
Total years in maintenance bilingual	0.04	0.00	0.06	0.04	0.18	0.17	3.43	2.91
Number of students	8,771	1,558	1,604	417	2,771	150	2,158	140

Note. CELDT scores are standardized by the total-sample standard deviation and centered at the respective cut-scores for test and year. CST scores are standardized and centered using state means and standard deviations by test, year, and grade. EL = English learner; IFEP = initially fluent English proficient; CST = California Standards Test; ELA = English language arts; CELDT = California English Language Development Test.

for their children on their school district intake form. All students are eligible for enrollment in general education/EI classrooms and DI classrooms. Only students who speak the target language at home are eligible for enrollment in TB and MB. Parents whose children are in one of the three two-language programs must sign a form each year indicating their program preference. This annual signature acts as a parent waiver allowing the district to offer two-language programming according to California law.

An important feature of this district is that kindergarten students are assigned to their instructional program and begin school before the district knows the results of their CELDT test and therefore prior to students' linguistic classification. Students can remain in their assigned program regardless of subsequent classification as EL or IFEP. As will be detailed later, most students who are enrolled in a two-language

program and find out that they are not considered as ELs (i.e., they are IFEPs) stay in their two-language program. Two-language programs are considered by many families to be desirable programs that support the maintenance and/or development of students' home languages. One consequence of this is that EL and IFEP students are in the same classrooms, with instructional differentiation between the two groups limited to ELD instruction for ELs, possible homogeneous in-class instructional group work, and one-on-one instruction.

Table 2 shows descriptive statistics of EL and IFEP students in each of the four instructional programs. Student characteristics vary meaningfully across and within programs. Both EL and IFEP students in the two bilingual programs are more likely to have characteristics that are associated with lower performance compared with EL and IFEP students in the EI and DI programs.

Comparing EL with IFEP students within programs, the table shows that the gap in relative academic performance (on second grade CST scores) is largest in the DI program and smallest in the TB program. Within each program, IFEP students have characteristics associated with higher performance compared with EL students.

Method

The study employs a binding-score RD with instrumental variables (IVs) design to examine the effects of EL classification on academic achievement trajectories. RD designs, when meeting the appropriate assumptions, offer rigorous causal estimates of treatment effects (Imbens & Lemieux, 2008). Standard regression estimates of EL status can be interpreted causally only if all observable and unobservable differences between EL and non-EL students are accounted for in the model. This is generally very difficult to do given that EL and non-EL students tend to differ in many ways. RD, by contrast, takes advantage of situations in which there is effectively random assignment of individuals into a treatment group (at the cusp) because treatment is assigned based on a set threshold on one or more continuous pretreatment covariates. In this case, students are assigned to EL status based on their CELDT scores. The premise of the method is that there is essentially random assignment of students to either the EL condition or the IFEP condition right at the cut-score. As will be shown, students who score one point lower on a CELDT subtest and who are classified as EL are no different, on average, than students who score one point higher and are classified as IFEP.

The trade-off for strong causal inference is that RD estimates only apply to students who are very close to the treatment threshold—in this case, this means language-minority students who enter kindergarten with relatively strong English proficiency levels. The results cannot be assumed to apply to students who enter kindergarten with low English proficiency levels. However, a significant number of ELs enter kindergarten with relatively strong English proficiency levels. In this sample, 18% of incoming language-minority students score within a quarter of a standard deviation (SD) above or below the EL–IFEP cut-score.

In effect, this RD estimates the difference between the CST scores of students who fall just below the EL–IFEP cut-score (and are classified as EL) and the CST scores of students who fall just above the EL–IFEP cut-score (and are classified as IFEP). RD uses data from students below the EL–IFEP cut-score to predict test score outcomes for students just below the cut-score. It uses data from students above the EL–IFEP cut-score to predict CST scores for students at or just above the cut-score. Any difference between estimated CST scores just below versus just at/above the cut-score is interpreted as the causal effect of EL versus IFEP status for students at the cusp of EL and IFEP classification.

To account for the longitudinal data used in this analysis, I embed the standard RD design in a growth model. The model, where i differentiates students, and t differentiates years in school, is described below.

Level 1: (1)

$$CST_{it} = \beta_{0i} + \beta_{1i}GRADE_{it} + e_{it}.$$

Level 2:

$$\begin{aligned} \beta_{0i} = & \gamma_{00} + \gamma_{01}CELDT_i + \gamma_{02}EL_i \\ & + \gamma_{03}CELDT_i EL_i + \Gamma_0 \mathbf{X}_i \\ & + \Lambda_0 \mathbf{PROGRAM}_i + \Upsilon_0 \mathbf{COHORT}_i + u_{0i}, \end{aligned}$$

$$\begin{aligned} \beta_{1i} = & \gamma_{10} + \gamma_{11}CELDT_i + \gamma_{12}EL_i \\ & + \gamma_{13}CELDT_i EL_i + \Gamma_1 \mathbf{X}_i \\ & + \Lambda_1 \mathbf{PROGRAM}_i + \Upsilon_1 \mathbf{COHORT}_i + u_{1i}, \end{aligned}$$

where $e_{it} \sim N(0, \sigma_e^2)$ and $\{u_{0i}, u_{1i}\} \sim N(\mathbf{0}, \Sigma)$,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}.$$

If a student scores below the cut-score on any of the required CELDT scores, he or she should be classified as EL. A student’s lowest score, therefore, can be thought of as a binding score. If the lowest score is at or above its cut-score, then the student should be classified as IFEP. If the lowest score is below its cut-score, the student should be classified as EL. I construct a rating variable, $CELDT_i$ in Equation 1, by taking each student’s lowest standardized initial CELDT score, centered at its respective cut-score. The rating variable is a continuous variable reflecting

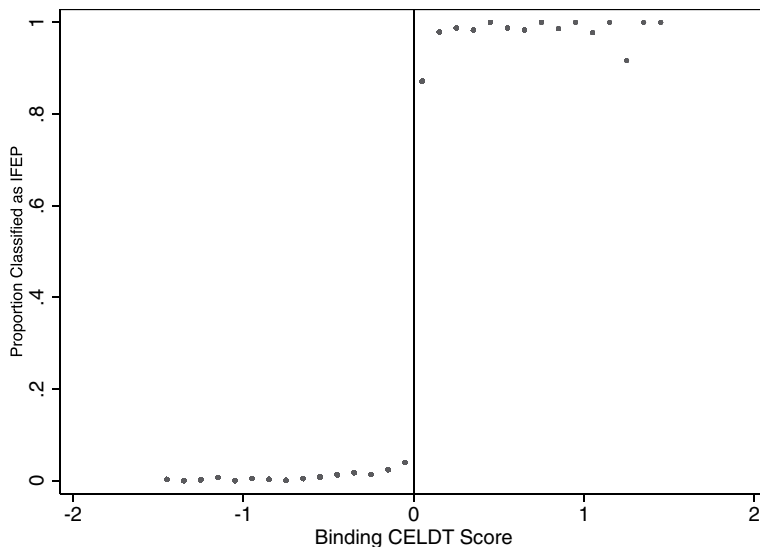


FIGURE 1. *Proportion of language-minority kindergartners classified as IFEP, by binding CELDT score.*
 Note. Binding CELDT score is a student’s lowest standardized CELDT score, centered at the IFEP cut-score. Bin size = .1. IFEP = initially fluent English proficient; CELDT = California English Language Development Test.

how far above or below the binding cut-score each student falls. This method is called binding-score RD and is described in Robinson (2011).

Level 1 represents how each student’s CST scores change linearly across grade. In Level 1, CST_{it} represents each student’s test score outcome, standardized using state means and standard deviations for each test in each year. CST tests are not vertically aligned and, therefore, are not ideal for growth modeling. But, by standardizing each student’s scores based on state means and standard deviations, I can compare scores across years (Valentino & Reardon, 2015). $GRADE_{it}$ is a student’s grade level, centered at Grade 2, the first grade in which students take the CST. Level 1 estimates are student specific, amounting to random effects for each student in the sample.

Level 2 of the equation represents how students’ CST scores differ based on EL or IFEP classification (Reardon & Robinson, 2012; Singer & Willett, 2003). γ_{01} and γ_{03} represent the estimated difference in second grade CST scores per unit change in the standardized CELDT variable, below and above the cut-score, respectively. γ_{11} and γ_{13} represent the estimated difference in the annual change in CST scores between Grades 2 and 7 (for math)/10 (for ELA) per unit change in the standardized CELDT variable, below and above the cut-score. I examined

plots and compared goodness of fit of different models to determine this linear functional form.

I use RD with IV (also called fuzzy RD) rather than RD alone because of imperfect compliance with school district EL and IFEP classification policies. In other words, not all students who should be classified as EL (or IFEP) based on their kindergarten CELDT score are classified as such. EL_i , the instrument in the model, is a dummy variable indicating whether the student should have been classified as an EL, based on his or her rating score, $CELDT_i$. To be a valid instrument, being eligible for EL status should (1) predict EL classification and (2) not directly impact CST scores. With regard to (1), compliance at the cusp with the district’s classification policy is high, approximately 89% (see Figure 1). In other words, 89% of students at the cusp are appropriately classified as EL or IFEP based on their CELDT score. Below, I describe how I account for imperfect compliance through use of the Wald estimator (Angrist, Imbens, & Rubin, 1996; Bloom, 1984). With regard to (2), I assume compliance with this assumption as it is very difficult to imagine any way in which hitting a certain CELDT score threshold would independently impact CST scores.

In the equation, X_i is a matrix of student background variables. It includes each student’s gender and ethnicity as well as if they were ever

identified for special education services and if they were born in the United States. Although inclusion of pretreatment covariates is not necessary for causal interpretations of RD estimates if all the appropriate model assumptions are met (see tests below), their inclusion can increase estimate precision and so I include them in the model. Unfortunately, I am unable to include information on students' economic background. The only variable the school district collects on this is eligibility for free or reduced priced lunch (FRPL) and these data are considered protected.

PROGRAM_{*i*} is a matrix of dummy variables, one for each initial (kindergarten) linguistic instructional program (EI is the omitted reference category). **COHORT_{*i*}** is a matrix of dummy variables, one for each cohort of students (with 2010 as the omitted category). Earlier cohorts are in the data for more years than more recent cohorts and, therefore, have more weight in the estimates. By including cohort fixed effects, I control for any average differences in the impact of EL classification on CST trajectories by cohort.

In this model, γ_{02} is the first parameter of interest, representing the average effect of EL status compared with IFEP status on marginal students' second grade CST scores (the first year that students take the CST). γ_{12} is the second parameter of interest. It represents the average effect of EL status compared with IFEP status on the incremental change in marginal students' CST scores by grade for each grade after second grade.

Once obtaining the parameter estimates, I test the joint significance of the two parameters of interest, γ_{02} and γ_{12} . Because the intercept and slope parameters are likely to be collinear, it is not sufficient to examine their individual *t* tests to determine whether or not EL students' trajectories, at the margin, differ from that of IFEPs. Instead, the joint significance test (an *F* test) tests whether the intercept and slope parameters together are statistically different from zero (Murray, 2006; Raudenbush & Bryk, 2002; Singer & Willett, 2003). A rejection of this null hypothesis indicates that there is a significant direct effect of EL classification on marginal students' academic trajectories, where trajectory does not refer solely to the slope parameter but rather is defined as students' second grade CST scores and annual change after second grade.

Although the parameter estimates for intercept and slope reveal the direction and magnitude of any effect of EL classification on the cusp, it is the test of joint significance that answers the research question posed here.

I run the models using a range of bandwidths of data on each side of the cut-score (0.25–1.25 SDs). A bandwidth of a given size uses that amount of data on the rating variable below and above the cut-score. The more data included, the more precise the estimate. However, including data that are too far from the cut-score can increase bias if the true functional form is not adequately modeled. I use the Imbens and Kalyanaraman (2012) method to calculate, for each CST outcome, the optimal bandwidth that balances precision with lack of bias. The optimal bandwidths cluster tightly around 0.75. Across analyses, the point estimates and statistical significance levels are very similar across bandwidths.

Equation 1 is an intent-to-treat model; it estimates the causal impact of EL classification on marginal students if there were 100% compliance. However, as stated above, 11% of students receive a language classification that does not reflect the appropriate classification based on their CELDT scores. To estimate the true effect of EL status on marginal student outcomes among those whose status complies with policy, I use the Wald estimator (Angrist et al., 1996; Bloom, 1984). The Wald estimator simply divides the point estimates and the standard errors of the two parameters of interest by the estimated effect of having met the criteria for EL classification on actually having been assigned EL status for students near the cut-score. Using the preferred bandwidth, the Wald estimator is 0.887.

Ideally, one would like to understand the impact of EL classification for all EL students. This study is only able to answer that question for students who enter kindergarten with relatively advanced English proficiency. Although this is a limitation of this article, this group of students is important from a policy and theory perspective. Students with little or no English ability have a pressing need for specialized services (Callahan et al., 2010). Students who are fully proficient in English have no need for specialized services. It is the students in the middle, with some English, but without full English proficiency, for whom policymakers struggle to determine and meet the needs. It is also these students who are most susceptible to being

TABLE 3

Check for Discontinuities at the EL–IFEP Cut-Score in Instructional Program Enrollment

	BW = 0.25	BW = 0.5	BW = 0.75	BW = 1	BW = 1.25
EI	0.00 (.04)	0.00 (.02)	0.00 (.02)	0.00 (.02)	0.01 (.02)
DI	–0.02 (.03)	–0.03 (.02)	–0.03[†] (.02)	–0.03* (.01)	–0.04** (.01)
TB	–0.01 (.02)	0.00 (.02)	0.00 (.01)	0.01 (.01)	0.01 (.01)
MB	0.04* (.02)	0.02 (.01)	0.02* (.01)	0.02* (.01)	0.02 [†] (.01)
<i>n</i>	2,754	5,694	8,756	11,139	12,710

Note. Estimates are of the impact of EL versus IFEP classification. Optimal bandwidth in bold. Robust standard errors appear in parentheses. EL = English learner; IFEP = initially fluent English proficient; BW = bandwidth; DI = dual immersion; TB = transitional bilingual; MB = maintenance bilingual; EI = English immersion.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

negatively impacted by EL status because they are less likely to reap strong positive benefits from EL programmatic offerings.

Model Checks. An assumption for RD is that the rating variable is smoothly distributed around the cut-score. I examined CELDT score distributions and found this to be the case (see Figure A in the appendix, available in the online version of the journal).

A second important assumption for RD is that no other factors impacting CST scores vary discontinuously at the EL–IFEP cut-score. If they do, then one cannot assign causal inference to the treatment variable. To test this, I examine whether there are any discontinuities at the cut-score for a range of covariates. Table B in the appendix (available in the online version of the journal) shows that there are no significant, meaningful discontinuities in covariates at the cut-score. The significance of this check is that it confirms that students just below the EL–IFEP classification are indistinguishable in observed characteristics, from those just above the EL–IFEP classification.

In addition, I conducted several tests to see whether the impact of EL classification varied significantly by student cohort. All the tests I conducted suggest that there are not significant differences in the EL effect by cohort. Finally, I conducted sensitivity checks on differential attrition or grade repetition at the margin. Results are discussed in the “Findings” section.

Linguistic Instructional Program Analysis. After exploring the main effect of EL classification on

students’ academic trajectories, this article goes on to examine how the effect of EL classification differs for students in different linguistic instructional programs. As described earlier, all students in this school district who speak a language other than English at home are assigned to programs prior to being classified as EL or IFEP. In effect, this means that students around the cut-score are essentially randomized into EL and IFEP status *within* each program.

This is akin to a blocked randomization design in a randomized controlled trial, where, for example, males and females are each randomly assigned to receive treatment. Students need not be randomly assigned to program and indeed they are not—Table 2 shows that students in the bilingual programs (TB and MB) tend to have characteristics correlated with lower linguistic and academic outcomes. The key is that within each program, students are essentially randomly assigned to EL or IFEP status at the cut-score. Because of this, I can estimate the effect of EL classification on marginal students in the four programs.

To test that instructional program assignment is independent of EL classification, I check for discontinuities in program enrollment at the cut-score. These results are in Table 3. There are no discontinuities in program enrollment in the two largest EL programs: EI and TB. There are small, marginally significant to significant, discontinuities in program enrollment in MB and DI. EL classification results in slightly more marginal students enrolling in MB and slightly fewer enrolling in DI. It is unclear how this occurs given that program assignment occurs prior to linguistic classification but conversations with district personnel suggest that a few students

may transfer programs if IFEP classification is conferred. This could be due to parental or administrative decisions. It is a small proportion of students suggesting little threat to causal estimates; however, estimates in these two programs (DI and MB) should be considered cautiously. Causal estimates in EI and TB remain rigorous.

The model interacts all noncontrol variables in Level 2 of Equation 1 with each of the three two-language instructional programs (EI is kept as the reference category). In effect, this calculates the effect of EL classification on marginal students separately for each program. The equation is described below.

Level 1: (2)

$$CST_{it} = \beta_{0i} + \beta_{1i} GRADE_{it} + e_{it}.$$

Level 2:

$$\begin{aligned} \beta_{0i} = & \gamma_{00} + \gamma_{01} CELDT_i + \gamma_{02} EL_i \\ & + \gamma_{03} CELDT_i EL_i + \Gamma_0 \mathbf{X}_i \\ & + \Lambda_0 \mathbf{PROGRAM}_i \\ & + H_0 CELDT_i \mathbf{PROGRAM}_i \\ & + M_0 EL_i \mathbf{PROGRAM}_i \\ & + \Psi_0 EL_i \mathbf{PROGRAM}_i CELDT_i \\ & + Y_0 COHORT_i + u_{0i}, \end{aligned}$$

$$\begin{aligned} \beta_{1i} = & \gamma_{10} + \gamma_{11} CELDT_i + \gamma_{12} EL_i \\ & + \gamma_{13} CELDT_i EL_i + \Gamma_1 \mathbf{X}_i \\ & + \Lambda_1 \mathbf{PROGRAM}_i \\ & + H_1 CELDT_i \mathbf{PROGRAM}_i \\ & + M_1 EL_i \mathbf{PROGRAM}_i \\ & + \Psi_1 EL_i \mathbf{PROGRAM}_i CELDT_i \\ & + Y_1 COHORT_i + u_{1i}, \end{aligned}$$

where $e_{it} \sim N(0, \sigma_e^2)$ and $\{u_{0i}, u_{1i}\} \sim N(\mathbf{0}, \Sigma)$,
 $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$.

As before, $\mathbf{PROGRAM}_i$ is a matrix of dummy variables for each of the three two-language programs: DI, TB, and MB. The estimates of interest here are γ_{02} and the vector of three estimates, \mathbf{M}_0 , which provide estimates of the impact of EL classification on the second grade CST scores of marginal students in each of the

four EL instructional programs, and γ_{12} and \mathbf{M}_1 , which provide estimates of the average effect of EL compared with IFEP status, by program, on the incremental change in marginal students' CST scores by grade for each grade after second grade. As in the main analysis, I compare goodness-of-fit to determine this functional form. \mathbf{M}_0 and \mathbf{M}_1 are estimates of the difference between each two-language program and EI's intercept, γ_{02} , and slope, γ_{12} , respectively.

To answer the research question posed here regarding whether the impact of EL classification differs across linguistic instructional programs, I conduct a series of F tests after running the models. The key test is the test of difference across linguistic programs. It tests whether the parameter estimates for the differences between EI and the two-language programs (slopes and intercepts) are equal to zero. If rejected, this test indicates that EL classification impacts students' trajectories differently in two-language programs compared with EI. Subsequently, I test whether the two-language program difference estimates for each program (slope and intercept) are equal to zero, allowing me to examine which specific two-language programs, if any, operate differently than EI. Finally, I test the joint significance of slope and intercept for each of the four programs. These tests examine whether there is an impact of EL classification on test score trajectories in each program.

Although this analysis provides causal estimates of the impact of EL classification on CST trajectories within each EL instructional program, one cannot assume that differences in EL classification effects across programs are directly due to treatments in those programs. This is because although EL and IFEP assignment is effectively random within each program, it is not random across programs. Students at the cusp of EL–IFEP classification in one program may differ from students at the cusp of EL–IFEP classification in another program. To the extent that EL classification impacts students differently in different programs, it may be due to program characteristics or student characteristics. To address this, I conduct a sensitivity analysis that includes fixed effects for parental school and program preference. This controls for many unobservable differences in students. I do not include these variables in the main analysis because I only have them for a subset of academic years. I discuss the results of this

TABLE 4

Estimates of Impact of EL Status on Math and ELA CST Scores, Binding-Score RD With IV Models

	BW = 0.25	BW = 0.5	BW = 0.75	BW = 1	BW = 1.25	Wald estimator
Math						
Intercept γ_{02}	-0.061 (.080)	-0.088 [†] (.047)	-0.073[†] (.042)	-0.066 [†] (.035)	-0.066* (.033)	-0.082[†] (.047)
Slope γ_{12}	-0.023 (.017)	-0.019 (.012)	-0.010 (.009)	-0.009 (.009)	-0.011 (.007)	-0.011 (.010)
Student covariates	X	X	X	X	X	X
Inst. program var.s	X	X	X	X	X	X
Cohort fixed effects	X	X	X	X	X	X
Joint sig. test (<i>p</i> value)	.131	.017*	.041*	.050 [†]	.014*	
<i>n</i> (student-semesters)	9,738	20,356	31,224	39,213	44,299	
ELA						
Intercept γ_{02}	-0.089 (.075)	-0.080 [†] (.046)	-0.054 (.037)	-0.059 [†] (.032)	-0.071* (.031)	-0.061 (.042)
Slope γ_{12}	0.004 (.013)	-0.009 (.007)	-0.008 (.006)	-0.007 (.005)	-0.006 (.005)	-0.009 (.007)
Student covariates	X	X	X	X	X	X
Inst. program var.s	X	X	X	X	X	X
Cohort fixed effects	X	X	X	X	X	X
Joint sig. test (<i>p</i> value)	.472	.024*	.045*	.033*	.014*	
<i>n</i> (student-semesters)	10,956	23,036	35,130	43,983	49,516	

Note. Intercept coefficients estimate the effect of EL status on second grade CST scores. Slope coefficients estimate the effect of EL status on the annual change in CST scores after the second grade. Joint significance tests examine whether the slope and intercept together are statistically significantly different from zero. Optimal bandwidth in bold. Robust standard errors appear in parentheses. Wald estimator is calculated using optimal bandwidth results. EL = English learner; ELA = English language arts; CST = California Standards Test; RD = regression discontinuity; IV = instrumental variable; BW = bandwidth; Instr. program var.s = instructional program variables; sig. = significance.

[†]*p* < .10. **p* < .05. ***p* < .01. ****p* < .001.

check and what it means for comparing estimates of the EL effect across programs. In addition, I conduct a sensitivity check to see whether EL marginal students, compared with IFEP marginal students, are more or less likely to remain in their initial program over time.

Findings

Net Impact of EL Classification

Point estimates suggest that there is a significant and growing negative effect of being classified as an EL as compared with an IFEP on CST math and ELA scores among students at the margin of EL–IFEP classification. Table 4 presents these results. The test of joint significance of the estimated effect of EL classification on second grade test scores (the intercept) and on annual test score change after second grade (the slope) at the optimal bandwidth is significant in both math and ELA, rejecting the null hypothesis that ELs and IFEPs at the margin have the same academic trajectories. As such, it is appropriate to examine the estimated intercept and slope coefficients, even if those coefficients are not statistically significant independently. That said, because

the individual parameter estimates are, by and large, not independently significant, I cannot reject the null hypothesis that the negative effect of EL status at the margin is limited to either the intercept or the slope parameters.

For math, EL students at the margin score 0.082 standard deviations lower than their IFEP counterparts in second grade (see “Wald estimator” column). The gap grows by 0.011 standard deviations in each subsequent year, reaching an estimated 0.139 standard deviations in seventh grade. This is a considerable effect size, amounting to 9.5 points on the CST by seventh grade, 27% of the statewide achievement gap between Latino and White students on the 2013 CST math test (California Department of Education, 2013).²

In ELA, the results are very similar and, like the results for math, the test of joint significance is significant at the .05 level. EL students at the margin score 0.061 standard deviations below their IFEP counterparts in second grade. The gap grows 0.009 standard deviations in each grade, reaching 0.133 standard deviations by the 10th grade. This is equivalent to 7.6 points on the CST, 22% of the Latino–White

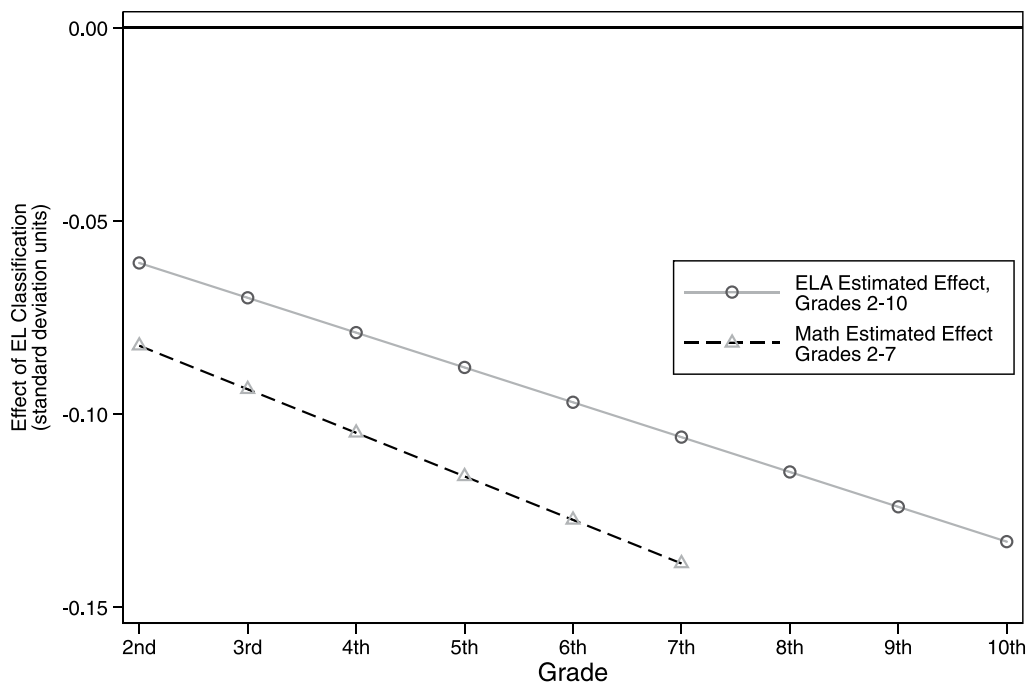


FIGURE 2. *Estimated effect of EL versus IFEP classification on ELA and Math CST test score trajectories.*
 Note. Bandwidth = 0.75 (Wald estimator). EL = English learner; IFEP = initially fluent English proficient; ELA = English language arts; CST = California Standards Test.

CST ELA statewide achievement gap (California Department of Education, 2013).

The EL effects for math and ELA from the preferred bandwidth are presented visually in Figure 2. This figure shows the estimated effect size of EL classification as the sample progresses through grade levels. Together, the math and ELA results tell a very consistent story. Among students who score just above or just below the EL cut-point on the CELDT when they enter the district in kindergarten, students do significantly and meaningfully worse on both math and ELA tests if they are classified as an EL rather than a fluent English speaker. Point estimates suggest that the penalty is meaningful in size by the second grade and grows slowly as students progress through school. The divergence in students' test scores is attributable to their classification as ELs and the bundle of treatments and services they receive as such.

Impact of EL Classification by Linguistic Instructional Program

Table 5 presents results and Figures 3 and 4 (for math and ELA, respectively) depict results

visually for the impact of EL versus IFEP status on marginal students' academic trajectories by instructional program. To conserve space, Table 5 shows only optimal bandwidth (0.75) results. Results from the full range of bandwidths are available in Table C in the appendix (available in the online version of the journal). These results show that the impact of EL classification on near-proficient students is not uniform; instead it operates differently in different instructional programs. The tests for both ELA and math show that the EL effect differs significantly between programs (i.e., difference between programs tests).

Although the effect of EL classification on marginal students in EI is significant and negative in both ELA and math, there is no significant effect (negative or positive) of EL classification in any of the other three programs. In EI, the EL penalty in second grade is sizable, amounting to a sixth of a standard deviation, and grows slowly across grades. Although sample sizes are smaller in the two-language programs, size alone is unlikely to explain differences between programs because point estimates suggest that EL classification may have a positive effect on marginal students in the

TABLE 5

Estimates of Impact of EL Status on Math and ELA CST Scores, by Initial Linguistic Instructional Program, Binding-Score RD With IV Model

	Math	ELA
EI intercept	-0.161*** (.041)	-0.143*** (.041)
DI intercept difference	0.145 (.113)	0.162 [†] (.089)
TB intercept difference	0.335** (.125)	0.378** (.134)
MB intercept difference	0.400* (.195)	0.315** (.102)
EI slope	-0.008 (.013)	-0.006 (.008)
DI slope difference	-0.021 (.029)	-0.004 (.019)
TB slope difference	-0.004 (.027)	-0.019 (.021)
MB slope difference	-0.012 (.044)	0.001 (.030)
Student covariates	X	X
Cohort fixed effects	X	X
<i>F</i> tests		
Difference between programs	.025*	.001***
DI different from EI	.401	.187
TB different from EI	.016*	.006**
MB different from EI	.111	.005**
EI joint test of significance	.000***	.000***
DI joint test of significance	.414	.816
TB joint test of significance	.344	.175
MB joint test of significance	.460	.166
<i>n</i>	31,224	35,130

Note. Estimates are from optimal bandwidth (0.75). Joint significance tests examine whether the slope and intercept together are statistically significantly different from zero. Difference tests examine whether EL classification effects are different between programs. Robust standard errors appear in parentheses. EL = English learner; ELA = English language arts; CST = California Standards Test; RD = regression discontinuity; IV = instrumental variable; EI = English immersion; DI = dual immersion; TB = transitional bilingual; MB = maintenance bilingual.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

TB and MB programs, and no effect on marginal students in the DI program. More specifically, I find that EL classification operates significantly differently in TB (math and ELA) and MB (ELA only) compared with EI. As a reminder, causal inference is strong in the EI and TB programs, but attenuated in MB and DI due to modest discontinuities in program enrollment at the EL–IFEP cut-score in those two smaller programs.

Sensitivity Checks

I conduct several sensitivity checks to probe the results of the net effect of EL classification and differences in the EL effect by linguistic program. First, I examine whether students on either side of the EL–IFEP cut-score were more or less likely to repeat a grade, exit the data, or otherwise not take the appropriate CST test in the

appropriate year. If EL-classified students are more or less likely to take the appropriate CST test, this could potentially explain the effect of EL classification on marginal students. Table D in the appendix (available in the online version of the journal) shows these results. There are no significant discontinuities in whether students on either side of the cut-point take the appropriate CST test in the appropriate year except for a small but significant discontinuity in the fifth grade ELA test. In that grade, EL-classified students at the margin are 2 percentage points more likely to take the appropriate CST test. It is unclear why this is but the small magnitude of the estimate and the localization to the fifth grade suggest that it does not explain the net effect of EL classification.

The second and third sensitivity checks pertain to the analysis by linguistic instructional program. As discussed above, the RD results are

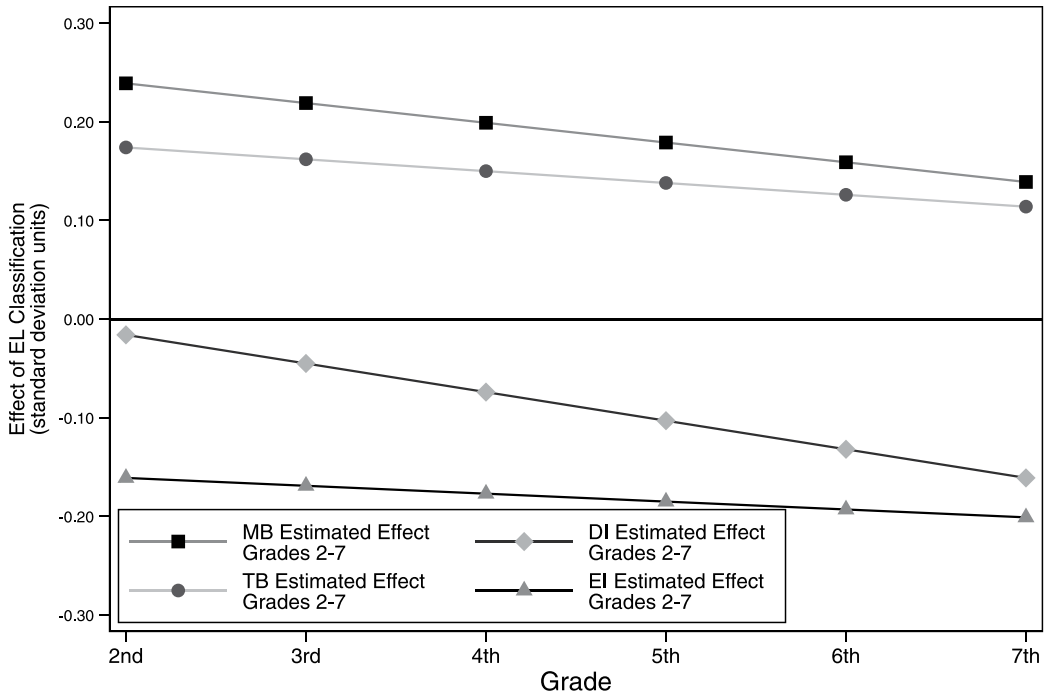


FIGURE 3. *Estimated effect of EL versus IFEP classification on Math CST test score trajectories, by linguistic instructional program and grade.*

Note. Bandwidth = 0.75. TB is significantly different from EI. MB and DI are not. EL = English learner; IFEP = initially fluent English proficient; CST = California Standards Test; TB = transitional bilingual; MB = maintenance bilingual; DI = dual immersion; EI = English immersion.

strong causal estimates within each instructional program, however to the extent that students at the cusp of EL–IFEP identification in one program are different from students at the cusp in another program, we cannot assume that we can compare the results across programs. One way to increase the comparability of effect estimates across programs is to control for parental preferences into programs. By controlling for parental preferences for school and instructional program (in addition to the other control variables in the model), I am able to control for many unobservable differences between students across programs. Therefore, the second sensitivity check includes fixed effects for parental choice (at Level 2) in the program-interacted model. In effect, this means that I am examining whether EL classification impacts academic outcomes among students at the cusp whose parents selected the same preferred school and program. Table E in the appendix (available in the online version of the journal) shows the results, which very closely parallel the model without parental

choice fixed effects. This finding suggests that one can cautiously compare EL effects across programs.

A final check on the instructional program results examines students' likelihood of remaining in their assigned program over time. If students on the cusp who are classified as IFEP are more likely to transfer out of the TB program and into the EI program, for instance, this may explain why EL classification is beneficial for students on the cusp in TB. In other words, EL-classified students on the cusp may get a fuller dose of a bilingual program than IFEP-classified students on the cusp. To analyze this, I examine the effect of EL versus IFEP classification on students' enrollment in each grade between first and fourth grades, among students at the cusp. Table F in the appendix (available in the online version of the journal) shows these results. In general, there is no evidence that classification at the cusp has a meaningful impact on perseverance in initial linguistic instructional program. In the first grade, EL classification

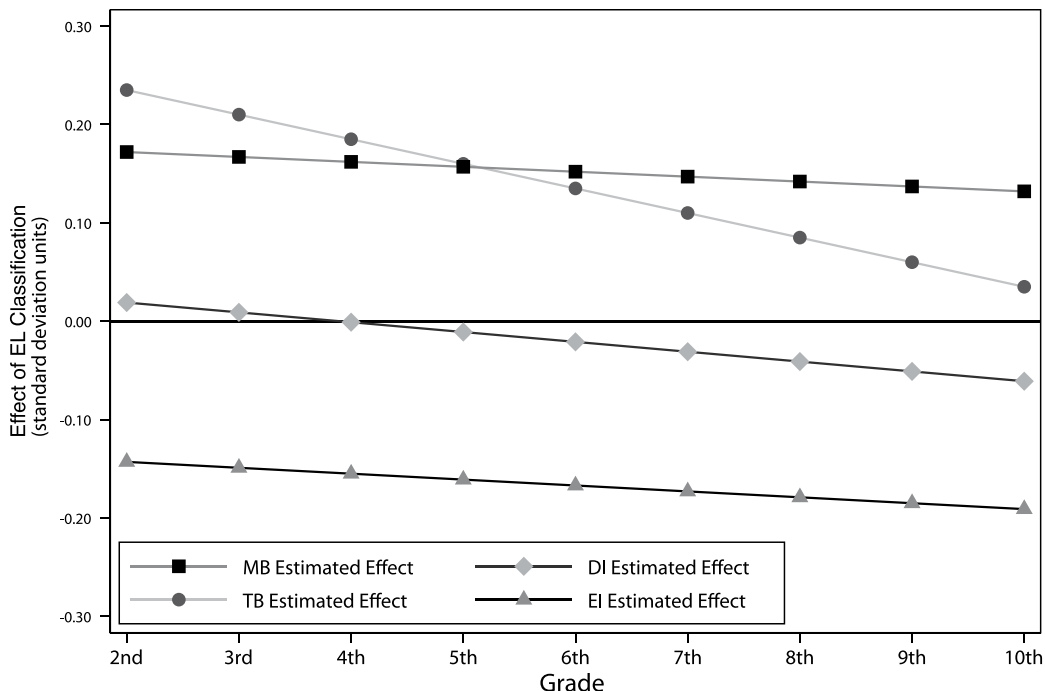


FIGURE 4. *Estimated effect of EL versus IFEP classification on ELA CST test score trajectories, by linguistic instructional program and grade.*

Note. Bandwidth = 0.75. TB and MB are significantly different from EI. DI is not. EL = English learner; IFEP = initially fluent English proficient; ELA = English language arts; CST = California Standards Test; TB = transitional bilingual; MB = maintenance bilingual; DI = dual immersion; EI = English immersion.

results in a 3-percentage point decline in DI enrollment and a 3-percentage point increase in MB. In the fourth grade, EL classification at the cusp results in a 4-percentage point increase in EI enrollment. The small magnitude of these discontinuities suggests that differential perseverance in program is unlikely to explain differences in the EL effect by initial instructional program.

Discussion

The current system of classifying students learning English is intended to avoid inequity in educational opportunity by offering students specialized services to meet their specific educational needs and by providing a mechanism for monitoring student progress and holding education systems accountable for that progress. Yet, this study finds that for some students, EL classification may in fact be contributing to educational inequity. For students in this district with relatively advanced English proficiency, EL classification in kindergarten results, on average, in a negative effect on academic achievement. Point

estimates suggest that this negative effect of EL classification is sizable and grows as students progress through school, although I cannot reject the null hypothesis that the negative effect is constant from second grade onward. A closer inspection, however, reveals that the negative effect is driven by the dominant instructional program in the district: EI. In two of the district's three two-language programs, results are quite different. Students at the cusp who start in TB and MB programs are not harmed by EL classification (in ELA and math in the transitional program and in ELA in the maintenance program).

The Negative Effect of EL Classification

Labeling theory provides a framework for understanding the finding that EL status, on average, hurts students at the cusp of EL-IFEP identification. Labeling theory suggests that students classified as ELs receive a bundle of treatments: These include both intended, programmatic services—targeted ELD instruction, language-accessible content instruction, specially trained teachers,

and annual assessments—and unintended treatments, including unintended status treatments or programmatic treatments (Link et al., 1989).

Robinson (2011) proposed that EL classification is likely to be beneficial for students with low English proficiency levels, harmful to students with high English proficiency levels, and neutral for students right at the point where they no longer require linguistic and academic supports. This set of hypotheses is a useful conceptual framework for understanding the likely effects of the intended programmatic treatments that come with EL classification: Because low English proficiency students need language supports more than higher proficiency students, the relative benefit of EL programmatic treatments will diminish as students acquire English.

However Robinson's framework does not take into account the status effects of EL classification.³ Diminished expectations on the part of teachers, internalization of negative stereotypes, and other status effects are likely to impact low and high English proficiency students alike. Students at the point at which they no longer benefit from specialized services will still suffer status loss and discrimination that arises from EL classification (Link & Phelan, 2013, Rist, 1977). This may explain why I find a negative net effect of EL status for students at the EL–IFEP cut-point in this district.

In addition, it is important to consider how EL programmatic services may, at times, be detrimental to students. This complicates the notion, in labeling theory, that the intended programmatic treatments that come with a label are likely to be beneficial to students whereas the unintended, status treatments are likely to be punitive (Link et al., 1989). Research on ELs not only confirms the negative status that often accompanies EL classification (Dabach, 2014; Thompson, 2015) but also suggests ways in which the very services that EL classification is designed to trigger may also penalize students. This may occur through inferior resource allocation in EL-specific classes (Gándara et al., 2003), EL services displacing academic content instruction (Estrada, 2014; Kanno & Kangas, 2014), or diminished rigor or coverage of academic content in EL classes (Dabach, 2014).

This, too, may explain the negative net effect of EL status for students at the cusp of EL–IFEP

classification in this district. Students with higher English proficiency levels benefit less from ELD and related services than those with more acute English language needs. But students with higher English proficiency remain vulnerable to negative effects of inferior resource allocation, diminished rigor, and displaced academic content instruction. Students at the cusp of EL–IFEP identification are the canary in the coalmine: With comparatively little to gain from EL programmatic services, they are a sensitive indicator of the effects of the problematic aspects of EL classification, both in terms of status effects and programmatic effects.

The results of this article suggest that the negative effects of EL status among marginal students in this district are evident by the second grade and may grow linearly over time after that. Although research has found negative consequences of remaining an EL in middle and high school (Callahan, 2005; Kanno & Kangas, 2014; Valenzuela, 1999), less research has examined the implications of EL classification in elementary school (with some exceptions such as Martin-Beltrán, 2010). The findings from this article, by contrast, suggest that some EL-classified students are negatively impacted by EL classification in elementary school. This finding does not undermine prior research on unique barriers to EL success in middle and high school, but it does focus attention on marginal ELs' experiences in elementary school.

The results of this study do not suggest that we should remove or hide the EL label from administrators, teachers, and peers because they apply only to students with relatively high English proficiency levels in kindergarten. Evidence suggests that lower proficiency students, as stated above, need and benefit from EL classification and treatments. Nor do the results suggest that we should simply lower the threshold for IFEP classification. This is because the impact of EL classification appears to depend, in part, not on the level at which classification is set, but on the services that EL-classified students receive.

Linguistic Instructional Program Moderates the Effect of EL Classification

The set of treatments that one student receives is not necessarily the same set of treatments that

another student receives. Indeed, treatments are likely to vary widely in different settings depending on district and school policies, individual student characteristics, and classroom and school practices and culture. Therefore, the negative net effect of EL classification found in this article should not be considered constant across districts, schools, or programs. Indeed, this study finds an important moderating factor: language of instruction.

This article finds that although the average net effect of EL treatments on near-proficient students is negative, the effect of EL classification in kindergarten is neutral for near-proficient students enrolled in TB (in math and ELA) and MB (in ELA). In other words, the negative effect of EL classification for marginal students is concentrated in EI classrooms. Bilingual classrooms, by contrast, may buffer students at the cusp from the negative treatments of EL status and/or bolster the positive ones.

Prior research suggests that both programmatic and status treatments triggered by EL classification are likely to differ between EI and two-language instructional environments, and these differences may result in differential impacts of EL classification. Programmatically, because EL students constitute the vast majority of students in most bilingual classrooms, teachers may be less likely to teach core academic content while ELs are pulled out for ELD instruction. This is in contrast to EI classrooms where EL students are often in the minority, and DI classrooms where EL students typically constitute half the class. If this scenario were accurate, ELs at the margin who are in bilingual (TB and MB) classrooms would be less likely to miss out on academic instruction and, as a result, may be less likely to fall academically behind their IFEP counterparts.

In terms of status effects, the social environment of two-language classrooms may be very different from monolingual classrooms. Rather than being defined by their lack of proficiency in one language, EL students in two-language classrooms may be defined by their knowledge of two languages. Indeed, bilingual education has at its root the development of positive intercultural relations (Gándara, 2005). In addition, teachers who speak the home language of their students may understand their

students' lives, backgrounds, and families better, and, as a result, have more accurate expectations of EL students and form closer relationships with them that help them succeed (Gifford & Valdés, 2006; Harklau, 1994; Stanton-Salazar, 1997). As a result, ELs in two-language programs may not experience stigma and discrimination at all or may not experience it to the same extent that students do in monolingual English programs.

It is interesting that the EL effect in DI, although not significant on its own, more closely parallels that of EI. This is notable given recent focus on DI as a promising alternative to both EI and traditional bilingual programs. This article raises questions regarding possible differences between DI and bilingual programs. As one example, although DI programs, like bilingual programs, focus on positive intercultural relations, some research suggests that there may be status hierarchies that favor English-dominant students within DI classrooms (Martin-Beltrán, 2010; Valdés, 1998a).

Although I find clear evidence that the impact of EL classification at the margin varies by linguistic program in this school district, it is important to remember that one should consider comparisons of results across programs cautiously. Students at the margin of EL-IFEP classification in bilingual programs are likely to be different, on observable and unobservable characteristics, from those at the margin in EI or DI. Simply moving EL students at the margin from immersion programs to bilingual programs may not resolve or reverse the EL penalty. This said, a sensitivity analysis failed to find evidence that differential effects of EL classification across programs are driven by differential selection into programs. In this sensitivity analysis, I included parental preferences for school and program, thereby controlling for many unobservable characteristics of students and their families (Valentino & Reardon, 2015). The results of this analysis were comparable with those of the main analysis, giving support for—but not conclusively affirming—the hypothesis that differences in treatments across programs drive differential effects of classification across programs.

Although RD analyses offer robust causal estimates when appropriate assumptions are met, a second caution with regard to these

results is the possibility of selection issues biasing results. Specifically, I found modest discontinuities in MB and DI enrollment at the EL–IFEP cut-score that introduce the possibility of some form of nonrandom selection into EL and IFEP classifications at the cut-score, at least within those two programs. That said, the strongest evidence that EL classification varies by program actually comes from the other two programs, EI and TB, minimizing concerns of bias.

Directions for Future Research

The findings from this study shift the discourse from locating the position of the “ideal” EL–IFEP cut-point (at the level in which students no longer need support services) toward a discussion of how to minimize the negative effects of EL classification on status and educational opportunity for students at all levels of English proficiency. To do so, we need a clear understanding of what the negative effects of EL classification are and how they operate.

In this article, I propose that EL classification is consequential to students due to both programmatic treatments and status treatments and that aspects of both types of treatments may penalize EL students, albeit often in unintended ways. We need research on both the programmatic and status mechanisms that operate to disadvantage EL-classified students at the cusp, particularly in EI classrooms. This will require both quantitative and qualitative research. In addition, we need to understand what causes the EL penalty to appear as early as the second grade, and why it may grow in magnitude from that point onward. Do the early effects set marginal students on a lower performing trajectory, and/or do students face compounding barriers to academic success as they move through school? The findings from this article also leave important questions as to how two-language classrooms, particularly TB and possibly also MB programs, buffer or even reverse the deleterious effects of EL classification and what other programmatic levers could potentially do the same. Finally, we need additional research on the impact of EL classification for students of less advanced English proficiency and in other districts and regions.

Conclusion

Court rulings and federal and state regulations have deemed language classification and specialized services necessary to ensure equality of educational opportunity for students learning the English language. Research suggests that these services can help students linguistically and academically. Yet, a growing body of work has identified numerous ways in which services for students learning English, and the classification of students by English language ability, creates a hierarchically tiered education system that parallels social inequalities outside of the educational setting (Callahan, 2005; Dabach, 2014; Dabach & Callahan, 2011; Gándara et al., 2003; Valdés, 2001).

Labeling theory (Link et al., 1989) provides a framework for analyzing this situation, suggesting that labels create opportunities through specialized services, and risks through stigma and discrimination. Together these opportunities and risks form a bundle of treatments that can be analyzed and adjusted. This article seeks to do just that. It asks, whether in one district and for students of relatively advanced English proficiency, the net impact of the bundle of treatments associated with EL classification is beneficial or harmful.

Among students with near English proficiency in kindergarten, this study finds evidence of a large, significant, and growing disadvantage among EL-classified students through the end of middle school in both math and ELA standardized tests. Yet, the effect of EL status is not monolithic: It depends on the services students receive. In particular, EL status may be neutral or even a boon to students enrolled in some bilingual classrooms.

Author’s Note

Any remaining errors are the sole responsibility of the author.

Acknowledgments

I am grateful to the following individuals for their thoughtful feedback and reflections: Sean Reardon, Kenji Hakuta, Claude Goldenberg, Tomás Jimenez, Martin Carnoy, Michelle Jackson, Aliya Saperstein, Rachel Valentino, Karen Thompson, Robert Linqanti, Joseph Robinson-Cimpian, Oded Gurantz, Eugene García, and Patricia Gándara. In addition, I extend

deep thanks to several individuals at the school district examined in this article. I will refrain from mentioning them by name to protect district anonymity, but these individuals provided invaluable insights into data features, program characteristics, and interpretation of findings.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded, in part, through two Institute of Education Sciences Grants, Award Numbers R305B090016 and R305A110670. It was also funded through the Spencer Foundation and National Academy of Education Dissertation Fellowship, and the Stanford University Graduate Student Fellowship.

Notes

1. There is some degree of variation in this process from state to state and district to district but the main components of this process are widely adhered to throughout the country (Abedi, 2008).

2. I compare the effect size with the Latino–White achievement gap rather than the EL–non EL achievement gap because the EL–non EL gap suffers from the revolving nature of EL classification (Saunders & Marcelletti, 2012). The effect size found in this article is also equivalent to a significant portion of the EL–non EL gap and the free and reduced priced lunch (FRPL)–non-FRPL CST gaps, as well as the Latino–White test score gap on the National Assessment of Educational Progress (NAEP; California Department of Education, 2013; Institute of Education Sciences, 2014). As a final point of comparison, the average effect size of an elementary school intervention on a broad standardized test score outcome like the CST is .07 (Hill, Bloom, Black, & Lipsey, 2008), roughly half the size of the EL penalty found here.

3. Much of the difference between Robinson's hypothesis and my own stems from the fact that we are asking different research questions. Robinson is asking about the effect of reclassifying students out of EL status and therefore compares current ELs with former ELs. The present article, by contrast, asks about the impact of EL classification and compares students who are classified as EL in kindergarten

(and often are reclassified later in elementary school) with students who never carry the EL classification. In terms of status effects, Robinson's entire sample has been vulnerable to the status effects of EL status, either currently or in the past, whereas my analysis compares students who are or have been vulnerable to status effects with those who have not. For this reason, status effects are much more likely to come into play in my analysis than in Robinson's.

References

- Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice*, 27(3), 17–31.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444–455.
- August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness*, 2, 345–376.
- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum.
- Bak, J. J., Cooper, E. M., Dobroth, K. M., & Siperstein, G. N. (1987). Special class placements as labels: Effects on children's attitudes toward learning handicapped peers. *Exceptional Children*, 54(2), 151–155.
- Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., . . . Linan-Thompson, S. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (IES Practice Guide, NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.
- Becker, H. (1963). *Outsiders: Studies in the sociology of deviance*. New York, NY: The Free Press of Glencoe.
- Bianco, M. (2005). The effects of disability labels on special education and general education teachers' referrals for gifted programs. *Learning Disability Quarterly*, 28, 285–293.
- Billings, E. S., Martin-Beltrán, M., & Hernandez, A. (2010). Beyond English development: Bilingual approaches to teaching immigrant students and English language learners. *Yearbook of the*

- National Society for the Study of Education*, 109, 384–413.
- Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8, 225–246.
- California Department of Education. (2013). *2013 STAR test results* [Database record]. Retrieved from <http://star.cde.ca.gov/star2013/SearchPanel.aspx>
- Callahan, R. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal*, 42, 305–328.
- Callahan, R., Wilkinson, L., & Muller, C. (2008). School context and the effect of ESL placement on Mexican-origin adolescents' achievement. *Social Science Quarterly*, 89, 177–198.
- Callahan, R., Wilkinson, L., & Muller, C. (2010). Academic achievement and course taking among language minority youth in US schools: Effects of ESL placement. *Educational Evaluation and Policy Analysis*, 32, 84–117.
- Callahan, R., Wilkinson, L., Muller, C., & Frisco, M. (2009). ESL placement and schools: Effects on immigrant achievement. *Educational Policy*, 23, 355–384.
- Carlson, D., & Knowles, J. (2016). The effect of English language learner reclassification on student ACT scores, high school graduation, and postsecondary enrollment: Regression discontinuity evidence from Wisconsin. *Journal of Policy Analysis and Management*, 35(3), 559–586.
- Chang, F., Crawford, G., Early, D., Bryant, D., Howes, C., Burchinal, M., . . . Pianta, R. (2007). Spanish-speaking children's social and language development in pre-kindergarten classrooms. *Early Education and Development*, 18, 243–269. doi:10.1080/10409280701282959
- Crawford, J. (1989). *Bilingual education: History, politics, theory, and practice*. Trenton, NJ: Crane Publishing Company.
- Dabach, D. (2014). "I am not a shelter!" Stigma and social boundaries in teachers' accounts of students' experience in separate "sheltered" English learner classrooms. *Journal of Education for Students Placed at Risk*, 19, 98–124.
- Dabach, D., & Callahan, R. (2011). Rights versus reality: The gap between civil rights and English learners' high school educational opportunities. *Teachers College Record*, ID Number: 16558. Available from <http://www.tcrecord.org>
- Domina, T., Penner, A., & Penner, E. (2014). *"Membership has its privileges": Student incentives and stigmatized identities in the accountability era*. Irvine: University of California.
- Estrada, P. (2014). English learner curricular streams in four middle schools: Triage in the trenches. *The Urban Review*, 46, 535–573.
- Gándara, P. (2005). Learning English in California: Guideposts for the nation. In M. M. Suarez-Orozco, C. Suarez-Orozco, & D. Qin (Eds.), *The new immigration: An interdisciplinary reader* (pp. 219–232). New York, NY: Brunner-Routledge.
- Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, 11, 1–54.
- Gifford, B., & Valdés, G. (2006). The linguistic isolation of Hispanic students in California. *Yearbook of the National Society for the Study of Education*, 105, 125–154.
- Goffman, E. (1963). *Stigma: Notes on a spoiled identity*. Englewood Cliffs, NJ: Prentice Hall.
- Goldenberg, C. (2013). Unlocking the research on English learners: What we know—and don't yet know—about effective instruction. *American Educator*, 37(2), 4–11, 38.
- Goldenberg, C., & Coleman, R. (2010). *Promoting academic achievement among English learners: A guide to the research*. Thousand Oaks, CA: SAGE.
- Gougeon, T. D. (1993). Urban schools and immigrant families: Teacher perspectives. *The Urban Review*, 25, 251–287. doi:10.1007/BF0111851
- Harklau, L. (1994). "Jumping tracks": How language-minority students negotiate evaluations of ability. *Anthropology & Education Quarterly*, 25, 347–363.
- Higgins, E. L., Raskind, M. H., Goldberg, R. J., & Herman, K. L. (2002). Stages of acceptance of a learning disability: The impact of labeling. *Learning Disability Quarterly*, 25, 3–18.
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2, 172–177.
- Imbens, G. W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79, 933–959.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142, 615–635.
- Institute of Education Sciences. (2014). *2013 mathematics and reading*. Washington, DC: Author.
- Jones, R. L. (1971). Labels and stigma in special education. *Exceptional Child*, 38, 553–564.
- Kanno, Y., & Kangas, S. (2014). "I'm not going to be, like, for the AP": English language learners'

- limited access to advanced college-preparatory courses in high school. *American Educational Research Journal*, 51, 848–878.
- Katz, S. (1999). Teaching in tensions: Latino immigrant youth, their teachers, and the structures of schooling. *The Teachers College Record*, 100, 809–840.
- Kim, J. S., Olson, C. B., Scarcella, R., Kramer, J., Pearson, M., van Dyk, D., . . . Land, R. E. (2011). A randomized experiment of a cognitive strategies approach to text-based analytical writing for mainstreamed Latino English language learners in grades 6 to 12. *Journal of Research on Educational Effectiveness*, 4, 231–263.
- Link, B., Cullen, F., Struening, E., Shrout, P., & Dohrenwend, B. (1989). A modified labeling theory approach to mental disorders: An empirical assessment. *American Sociological Review*, 54, 400–423.
- Link, B., & Phelan, J. (2013). Labeling and stigma. In C. S. Aneshensel, J. C. Phelan, & A. Bierman (Eds.), *Handbook of the sociology of mental health* (pp. 525–541). Dordrecht, The Netherlands: Springer.
- Martin-Beltrán, M. (2010). Positioning proficiency: How students and teachers (de)construct language proficiency at school. *Linguistics and Education*, 21, 257–281.
- Master, B., Loeb, S., Whitney, C., & Wyckoff, J. (2012). *Different skills? Identifying differentially effective teachers of English language learners* (Working Paper No. 68). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.
- McDermott, R. (1993). The acquisition of a child by a learning disability. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 269–305). Cambridge, UK: Cambridge University Press.
- Morgan, P. L., Frisco, M. L., Farkas, G., & Hibel, J. (2010). A propensity score matching analysis of the effects of special education services. *The Journal of Special Education*, 43, 236–254.
- Murray, M. P. (2006). *Econometrics: A modern introduction* (Vol. 929). Boston, MA: Pearson Addison Wesley.
- Papay, J. P., Murnane, R. J., & Willett, J. B. (2010). The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis*, 32, 5–23.
- Pennycook, A. (2002). *English and the discourses of colonialism*. London, England: Routledge.
- Ramsey, A., & O'Day, J. (2010). *Title III policy: State of the states*. Washington, DC: U.S. Department of Education.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: SAGE.
- Reardon, S., & Robinson, J. (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, 5, 83–104. doi:10.1080/19345747.2011.609583
- Rist, R. C. (1977). On understanding the processes of schooling: The contributions of labeling theory. In J. Karabel & A. H. Halsey (Eds.), *Power and ideology in education* (pp. 292–305). New York, NY: Oxford University Press.
- Robinson, J. (2011). Evaluating criteria for English learner reclassification: A causal-effects approach using a binding-score regression discontinuity design with instrumental variables. *Educational Evaluation and Policy Analysis*, 33, 267–292.
- Robinson-Cimpian, J., & Thompson, K. D. (2014). The effects of changing test-based policies for reclassifying English Learners: A difference-in-regression-discontinuities approach. *Journal of Policy Analysis and Management*, 35, 279–305.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3, 16–20.
- Saunders, W., Foorman, B., & Carlson, C. (2006). Is a separate block of time for oral English language development in programs for English learners needed? *The Elementary School Journal*, 107, 181–198.
- Saunders, W., Goldenberg, C., & Marcelletti, D. (2013). English language development: Guidelines for instruction. *American Educator*, 37, 13–25.
- Saunders, W., & Marcelletti, D. (2012). The gap that can't go away: The catch-22 of reclassification in monitoring the progress of English learners. *Educational Evaluation and Policy Analysis*, 35, 139–156.
- Scheff, T. J. (2009). *Being mentally ill: A sociological theory* (3rd ed.). New Brunswick, NJ: Transaction Books.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.
- Slavin, R., Madden, N., Calderón, M., Chamberlain, A., & Hennessy, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis*, 33, 47–58.
- Spack, R. (1997). The rhetorical construction of multilingual students. *TESOL Quarterly*, 31, 765–774.
- Stanton-Salazar, R. D. (1997). A social capital framework for understanding the socialization of racial minority children and youths. *Harvard Educational Review*, 67, 1–41.

- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629.
- Sullivan, A. L., & Field, S. (2013). Do preschool special education services make a difference in kindergarten reading and mathematics skills? A propensity score weighting analysis. *Journal of School Psychology, 51*, 243–260.
- Thomas, W., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Berkeley, CA: Center for Research on Education, Diversity & Excellence.
- Thompson, K. D. (2015). Questioning the long-term English learner label: How classification and categorization can blind us to students' abilities. *Teachers College Record, 117*, 1–50.
- Umansky, I. (2015, April). *The impact of English learner state on academic course-taking*. Paper presented at the American Educational Research Association, Chicago, IL.
- Valdés, G. (1998a). Dual-language immersion programs: A cautionary note concerning the education of language minority students. *Harvard Educational Review, 67*, 391–430.
- Valdés, G. (1998b). The world outside and inside schools: Language and immigrant children. *Educational Researcher, 27*(6), 4–18.
- Valdés, G. (2001). *Learning and not learning English: Latino students in American schools*. New York, NY: Teachers College Press.
- Valentino, R., & Reardon, S. (2015). Effectiveness of four instructional programs designed to serve English language learners: Variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis*. Advance online publication. doi:10.3102/0162373715573310
- Valenzuela, A. (1999). *Subtractive schooling: US-Mexican youth and the politics of caring* (SUNY Series, The Social Context of Education). Albany: State University of New York Press.
- Vollmer, G. (2000). Praise and stigma: Teachers' constructions of the "typical ESL student." *Journal of Intercultural Studies, 21*, 53–66.
- Walqui, A. (2006). Scaffolding instruction for English language learners: A conceptual framework. *International Journal of Bilingual Education and Bilingualism, 9*, 159–180.

Author

ILANA M. UMANSKY is an assistant professor in the College of Education at the University of Oregon. Her research uses quantitative methods and sociological theory to examine the educational opportunities and outcomes of immigrant students and students classified as English learners.

Manuscript received April 16, 2015

First revision received January 25, 2016

Second revision received June 3, 2016

Accepted July 25, 2016