NBER WORKING PAPER SERIES

EFFECTS OF ENGLISH LEARNER RECLASSIFICATION POLICIES ON ACADEMIC
TRAJECTORIES

Julian Betts
Laura Hill
Karen Bachofer
Joseph Hayes
Andrew Lee
Andrew Zau

Effects of English Learner Reclassification Policies on Academic Trajectories
Julian Betts, Laura Hill, Karen Bachofer, Joseph Hayes, Andrew Lee, and Andrew Zau
NBER Working Paper No. 28188
December 2020
JEL No. I21,I24,J1,J15

ABSTRACT

In California 38% of students are either current or former English Learners (ELs). A crucial decision in their educational trajectory is when to reclassify ELs. Upon reclassification, ELs cease to receive language supports, but have more opportunities to take the same courses as fluent English speakers. This paper uses regression discontinuity in California's two largest school districts to ask: "Are ELs being reclassified at the right time?" In most cases the districts were reclassifying students appropriately, with no discontinuity for outcomes including test scores and graduation on time. Exceptions in which those meeting the reclassification criteria subsequently performed worse were resolved with Los Angeles's newer reclassification policy.

Julian Betts
Department of Economics, 0508
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
and NBER
jbetts@ucsd.edu

Laura Hill
Public Policy Institute of California
500 Washington Street, Suite 600
San Francisco, CA 94111
hill@ppic.org

Karen Bachofer
Department of Economics, UCSD
9500 Gilman Drive
La Jolla, CA 92093-0508
karenbachofer@gmail.com

Joseph Hayes
Public Policy Institute of California
500 Washington Street, Suite 600
San Francisco, CA 94111
hayes@ppic.org

Andrew Lee
Public Policy Institute of California
500 Washington Street, Suite 600
San Francisco, CA 94111
andrew.lee.425@gmail.com

Andrew Zau
Department of Economics
UCSD
9500 Gilman Drive
La Jolla, CA 92093-0508
azau@ucsd.edu

Introduction

Many government programs seek to provide temporary assistance to people with special needs. In education, two examples include additional supports to students receiving special education services, and supports given to English Learners. In addition, there are many federal- and state-funded interventions designed to provide temporary assistance to struggling schools or students. Examples of temporary assistance from policy areas outside education include unemployment insurance for unemployed workers, and welfare support in general in the U.S. A key question in all of these cases is "When should the supports be removed, if at all?". While it is easy to study those for whom a support service is removed, it is not obvious how to create the counterfactual outcome, that is, what would have happened to this student or this worker had the support continued? By choosing a comparison group, the researcher creates the risk of a biased analysis due to unobserved differences between the people from whom services were removed and the people in the comparison group. An increasingly used solution to this problem is to adopt a regression discontinuity approach, where the comparison group consists of students or adults who were almost eligible to have a service removed, based on a continuous variable such as a test score.

In this paper we address a problem of this nature related to language services that schools provide to the many English Learner students in the United States, to help them master English. English Learner (EL) status is meant to be temporary—when students demonstrate sufficient English language proficiency, their official designation changes to Reclassified Fluent English Proficient (RFEP). If ELs are reclassified too soon, their academic performance may falter because their lack of mastery in English may impede their understanding of course materials. But if ELs are reclassified too late, their academic progress may also falter. In this case the language supports the ELs are receiving could create an opportunity cost in the form of reduced opportunities to take the same courses as native English speakers. This paper uses student-level data secured though special arrangements with the two largest school districts in California to ask "Are ELs being reclassified at the right time?"

It is important to study reclassification because it touches upon the lives of so many students. In total, more than 40 percent of the students in California's public schools speak a language other than English at home. In the 2016–17 school year, 21 percent (or more than 1.3 million) of all students were English Learners. When students who were formerly English Learners are added in, the population of "ever ELs" expands to 38 percent of all K–12 students in the state. Nationwide, 10 percent of all K-12 students are English Learners.[1]

A second reason reclassification is important concerns the recent passage of the federal Every Student Succeeds Act (ESSA). ESSA now requires states to standardize reclassification policies across all of their school districts. States such as California that allow their criteria to vary across districts will need to align districts to one state standard. Massachusetts and Florida are two other relatively diverse states that also allow districts to adapt state guidance on reclassification.[2]

---

[1] National Center for Education Statistics (2019). English Language Learners in Public Schools.
[2] States where districts follow a singular state policy may have districts that implement it in different ways, either by applying different weights to each of the reclassification criteria, or having subjective criteria, such as teacher input. See Mavrogordato

In addition, California has recently introduced two new tests to assess the overall English language arts proficiency of all students and the language mastery of ELs specifically. We must understand the impact of reclassification policies that relied on the old tests in order to create a roadmap for developing the new reclassification policies.

Students entering K–12 schools in California are classified as English Learners if they speak a language other than English at home and score below a proficiency threshold on the California English Language Development Text (CELDT).[3] English Learners are meant to have English language development instruction, either in stand-alone English language development classes ("designated") or as part of regular instruction ("integrated")[4]. EL students do receive core subject instruction, but English language development instruction may mean they have less of it, especially in high school (Callahan 2005; Kanno and Kangas 2014; Parrish et al. 2006), if for example, an ELD course fills a slot in a students' schedule that could otherwise be used for additional core instruction. After they are reclassified, these students no longer receive English language development instruction and take core subjects without additional support.

In order to be reclassified, students must demonstrate English language proficiency on assessments administered only to EL students and demonstrate basic skills in English that are comparable to native English speakers. Individual school districts make the decision to reclassify EL students based on a number of criteria recommended by the state.

Student outcome data have long revealed that EL students do not fare as well as their non-EL peers. However, reclassified students perform much better than current EL students, and they sometimes perform better than students who do not speak a foreign language at home. (See e.g. Hill, Weston, and Hayes, 2014; Hill et al., 2014; Saunders and Marcelletti,2013; Gándara and Rumberger, 2006.) Yet this does not imply any causal effect of reclassification. Without clarity about causality, it is difficult to design reclassification policies that optimize the duration of English language support.

This paper uses student-level data in California's two largest school districts (Los Angeles Unified and San Diego Unified), to evaluate the effect of reclassification on academic outcomes for former ELs. [5] These two large districts account for 14% of ELs statewide and 4% of ELs nationwide.

We use a Regression Discontinuity design (RD) to examine outcomes for students just above and just below the cutpoints on various tests used to make reclassification decisions. Because the two districts we study have used four different reclassification policies over the period we consider, we can identify not only any causal relationship between duration of English language support

and student outcomes, but also which of the reclassification standards comes closest to reclassifying a student at the appropriate time. If a policy reclassifies a student too soon, we detect a positive effect of an extra year of EL support. If a policy reclassifies too late, we detect a negative effect. And if we found no significant difference, it would suggest that students are being reclassified appropriately, so no change is optimal.

This is an important time to be thinking about how we measure when ELs are prepared to fully integrate into academic courses without English language support. Nationwide, implementation of the common core standards and new English language development standards are underway, but it is uneven across California (Warren and Murphy 2014; McLaughlin, Glaab and Carrasco 2014). The new statewide standards test, the Smarter Balanced Assessment Consortium's (SBAC) test, was administered for the first time in 2014-15, and the test results for the first year were particularly alarming for EL students (Hill and Ugo, 2016). The replacement for the CELDT, the English Language Proficiency Assessment for California (ELPAC), was fully implemented in the 2017-18 school year. Further, some research suggests it may be beneficial to use just one criterion, such as the ELPAC, to assess whether EL students are ready for reclassification (Umansky et al. 2015 and Hill et al. 2014).

The next section provides background on reclassification policies and explains what circumstances could lead to better or worse outcomes. The following sections review the literature, data, methods, and results.

Background on English Learner Reclassification Policies

We focus on EL students and reclassification decisions made in LAUSD and SDUSD between 2002 and 2012.[6] In California, the state issues guidance about reclassification policy, and district staff (EL program administrators and teachers) may modify it and implement reclassification decisions.[7] The policies and practices for identifying EL students, assigning them to instructional programs, and reclassifying them are clearly articulated in each school district's *Master Plan for English Learners*. Although policies vary somewhat between the districts, the policies are guided by state law, the California Department of Education, and the State Board of Education.

California's reclassification guidelines require the use of four criteria (California Education Code, Section 313(f)). During the period we study, the reclassification guidelines included:

1. The California English Language Development Test (CELDT), with recommended overall and subtest scores;
2. A test of basic skills in English, with a recommendation of the California Standards Test (CST) of English Language Arts (ELA) and a minimum score on the test;
3. Teacher evaluation; and
4. Parent consultation.

School districts in California, as in many other states (cited in Kim and Herman, 2012), are allowed to determine their own reclassification policies as long as they follow minimum suggested guidelines issued by the state. However, most states (30) in the pre-ESSA era relied only on an English Language proficiency assessment (Linquanti and Cook 2015). In California before January 2019, districts decided how to implement the four state reclassification criteria by

---

6 Through 2014 in SDUSD.
7 Decision-making and implementation are the responsibility of the Office of Language Acquisition in SDUSD and the Multilingual and Multicultural Education Department in LAUSD.

setting local policy, but since January 2019, the English proficiency criteria have been standardized statewide.[8] The policies and practices for identifying EL students, assigning them to instructional programs, and reclassifying them are articulated in each school district's *Master Plan for English Learners*.[9]

A 2013 survey found most California school districts had developed more rigorous reclassification standards than those recommended by the state (Hill, Weston, and Hayes, 2014). For example, many districts required higher cut scores on the CST ELA or on the CELDT, required the CST math as another measure of basic skills, or required course marks for teacher evaluation. Results from the survey suggest parental consultation was of limited importance.

Like most other California districts, SDUSD and LAUSD both have more rigorous standards than the minimum state guidelines, but the rigor of their policies differs, providing a unique opportunity to examine the relationship of more and less rigorous criteria to reclassified students' outcomes (*Master Plan,* LAUSD 2012 and SDUSD 2009). Earlier research by the paper's authors found that more rigorous criteria in SDUSD and LAUSD are associated with academic outcomes for reclassified students that are roughly the same as, or better than, those of native English speakers, respectively. LAUSD had more stringent criteria than SDUSD during the period we studied, and this was associated with lower reclassification rates and stronger outcomes for reclassified students in Los Angeles, relative to San Diego (Hill et al., 2014). However, these results could derive from compositional effects – causal effects were not investigated.

The Los Angeles and San Diego Unified School Districts (LAUSD and SDUSD) have had their English language instructional programs in place for well over a decade, and the policies for determining reclassification in both districts have remained the same since 2006 through the final year in which we study reclassification, 2012 in LAUSD and 2014 in SDUSD. Different policies were in place in both districts from 2002-2006. For example, SDUSD raised its basic skills reclassification criterion for students at all grade levels in 2005-06 and then lowered the threshold to the current level for secondary students in 2006-07. Similarly, the LAUSD criterion for teacher evaluation of EL students specified minimum mathematics marks from 2002-03 to 2005-06, but subsequently dropped the requirement that ELs meet a mathematics mark threshold. Both districts were subject to the state's rescaling of its language test (which made it more difficult) in the 2006-07 school year.

Why Reclassification Could Produce Better Student Outcomes

There is a clear rationale for providing additional language support for English Learners. Classrooms designed for native English speakers are designed with the presumption that students already have a certain ability to comprehend and speak English. A further presumption is that students are roughly at grade level in reading and writing. The basis for EL support programs is that the rate of language acquisition is assumed to be slower for an EL student placed in a traditional class than for one placed into a special class with additional language supports because facility with English is a prerequisite for understanding teachers and curricular

---

[8] CDE's current guidance, as of January 2019, is found in the Updated Reclassification Guidance for 2018-19. The new guidance specifies that an ELPAC Overall Performance Level (PL) 4 is required for reclassification. In order to assess whether an EL meets the ELA basic skills requirement, districts may use either a local assessment or the Smarter Balanced Summative Assessment (grades 3-8, 11).

[9] Decision-making and implementation are the responsibility of the Office of Language Acquisition in SDUSD and the Multilingual and Multicultural Education Department in LAUSD.

materials. However, as ELs' proficiency with English grows, at some juncture they may actually perform as well or better in classes which native English speakers typically take.

Figure 1 illustrates hypothetical annual test score gains depending on whether a student receives EL support. The blue line shows gains for the student if the student receives EL support in all years. The brown line shows the gains if the student does not get EL support. To the left of the vertical line, a student needs EL support to maximize test score gains. To the right of the vertical line, a student should be reclassified to maximize test score gains. Where the blue and brown lines intersect, a hypothetical student should transition from receiving EL services to being reclassified because annual test score gains become identical whether or not students get EL support. Thus, the optimal time to reclassify a student is shown by the vertical black line. In our experiments, if we see that students see a dip in their test scores post reclassification, this suggests that they were reclassified too soon and that reclassification criteria were too loose. If we observe that test scores jump post reclassification, this indicates students were reclassified later than would have been optimal, possibly holding them back from mainstream instruction or additional academic content. If we find no statistically significant difference in test score gain post reclassification, then reclassification criteria were optimally set for student progress.

Figure 1 Test Score Gains Rise, Fall, Or Stay The Same Depending On Whether Reclassification Occurs At The Right Time
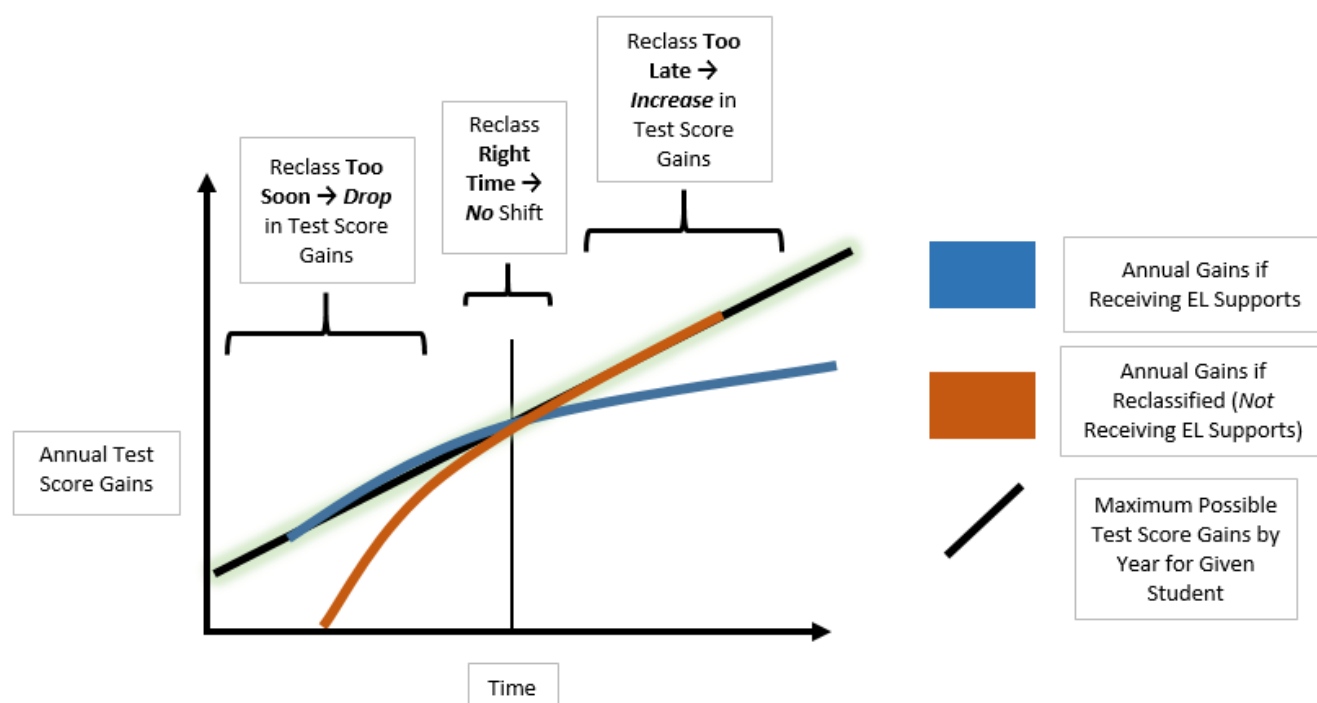
SOURCE: Authors.
NOTE: Hypothetical effects of EL reclassification

Relevant Literature

Ideally, students are reclassified at the moment when EL support no longer benefits them and they are prepared to undertake an English-only instructional program without that support. Because reclassification policies vary widely, it is unlikely that all school district policies are pegged to the moment of most benefit for an individual student. Capitalizing on our ability to precisely identify the reclassification policies used to decide when EL support should be removed, we estimate the causal relationship between continued EL instructional support and academic outcomes for students who are performing near reclassification cutoffs under four different reclassification policies (covering two policy eras in each district).

How does being reclassified affect student outcomes? Prior research has found that reclassified ELs are among the best performing students on a variety of academic measures (Hill et al., 2014; Saunders and Marcelletti, 2013; Gándara and Rumberger, 2006; EdSource, 2008; Flores, Painter, and Pachon, 2009). And prior research focusing on the rigor of reclassification criteria finds that students reclassified under more rigorous criteria in elementary school grades usually have slightly better outcomes than those reclassified under less rigorous criteria (Hill et al., 2014; Kim and Herman, 2012), but that those reclassified at older ages with more rigorous criteria may not (Hill, Weston, and Hayes, 2014). Again, these are observational studies and say nothing about causation. The mere act of increasing standards for reclassification should mechanically improve outcomes for reclassified students by removing the left tail of the original distribution of reclassified students.

We hypothesize that the decision to reclassify could influence shorter-term academic outcomes, as well as longer-term outcomes such as graduating on time. There are several moderating factors that we will ultimately consider in addition to the rigor of the reclassification criteria, which can vary across districts (Parrish et al., 2006; Hill, Weston, and Hayes, 2014). For example, prior research has demonstrated that grade at reclassification is an important correlate of outcomes (Hill, Weston, and Hayes, 2014; Hill et al., 2014).

Ours is one of several studies that have had access to the student-level data needed to establish causality using and RD approach. We use the data to examine whether reclassified students do better or worse *as a result of reclassification* than their counterparts who just miss the reclassification cutoff. The present paper is the first to ask this question in San Diego, and the third to study the question in Los Angeles. Pope (2016), examining LAUSD reclassification decisions from 2002–03 to 2003–04, looked at the impact of being just above or below the CELDT overall cutpoint.[10] Our study covers a much longer period and examines a larger set of the reclassification criteria. Robinson-Cimpian and Thompson (2016) investigated two LAUSD reclassification policy eras. This study looked at only one of five reclassification criteria (the CST, which measures ELA basic skills), and therefore did not fully evaluate whether the overall policy in these periods was set appropriately.[11]

---

[10] Pope found that 2nd and 4th grade students just above the CELDT overall cutpoint had improved English test scores and GPA relative to those just below, and there were but no statistically significant differences for high school students just above or below the CELDT overall cutpoint.

[11] We highlight here the similarities and differences between our report and the Robinson-Cimpian and Thompson paper (2016), because its focus is LAUSD and for similar school years. This important paper focuses on the CST ELA RD experiment, but uses fewer years (2002–03 to 2009–10) than our current study. In addition, our paper is different in that we separate elementary

Other recent papers ask similar questions in different school districts. Robinson (2011) studied reclassification policy at a single point in time in an unnamed medium-sized California district, finding that the district had appropriate reclassification criteria for elementary and middle school students, but may have reclassified high school students too soon. Carlson and Knowles (2016) use comprehensive administrative data from the Wisconsin Department of Public Instruction from 2006-07 to 2012-13 to estimate the effects of reclassification—specifically in 10th grade—upon subsequent outcomes. The authors report positive and significant effects on each of the outcomes they examined: ACT scores, graduation from high school, and enrolling at a postsecondary institution directly after graduation.

Cimpian et al. (2017) present quite a different context—they use longitudinal data on two separate states' EL students to examine the effect of reclassification upon subsequent achievement, while comparing districts. They present results first for their estimates of statewide average effects, and then for inter-district variability within each state. The state-level analysis yields negligible estimates of reclassification's benefits to subsequent achievement. But intra-state variation in district policies is large; in some districts, they find a large, significant negative effect upon graduation; in others, a large and significant positive effect. Like our current paper, Cimpian et al. demonstrate the importance of examining multiple reclassification policies and educational contexts.

Reyes and Hwang (2019) use an RD approach to examine a specific event—reclassification by the end of 8th grade in an unnamed southern Californian district. Non-causal results indicated that reclassification was associated with higher standardized test scores and better behavioral outcomes. But using the RD approach they find null effects of reclassification on each of the subsequent outcomes examined: CST-ELA scores, results on the California High School Exit Examination (CAHSEE), math course placement in high school, absences, and suspensions. The authors conclude reclassification thresholds in their district are appropriately assigned.

Johnson (2020) also uses a regression discontinuity approach to examine outcomes for students in a large unnamed California district during a time frame similar to ours. In this case, the event under consideration is reclassification in the 8th grade. Johnson is similarly unable to reject the null hypothesis of no effect of 8th grade reclassification upon all three outcomes: ELA scores in 9th, 10th, and 11th grades; SAT reading scores; and being on-track to graduate in 10th and 11th grades.

Data and Measures

Our sample includes all EL students in grades 3 through 12 in LAUSD and SDUSD for the school years between 2001-02 and 2015-16. Because our RD design will control for baseline characteristics from the year before reclassification, with outcomes measured one or more years after reclassification, we require a student to be in the district for three consecutive years.

and middle school students, running separate experiments for elementary, middle, and high school grades. The Robinson-Cimpian and Thompson paper focuses on only Latino students, while our analyses include EL students of all ethnicities, while controlling for whether a student's home language is Spanish. Probably the most important difference is that we also account for the teacher recommendation component of LAUSD's reclassification policy (i.e. the requirement of math and ELA course marks through 2005--06 and ELA course marks from 2006–07 onwards). Our study also goes beyond the focus on CST requirements by examining two elements of the CELDT scores to assess whether those cutoffs were appropriate. Finally, as already mentioned, our report includes another important school district: SDUSD.

Our outcomes include standardized test scores, which are available for students in multiple grades, and high school graduation. Because the goal of EL programs is to accelerate language ability of EL students until they have reached the point where they can perform well in classes designed for native English speakers, some of our main outcomes relate to English Language Arts (ELA) performance. Our chief measure of ELA skills is the grade-specific CST ELA test. We convert the CST scores in years after the reclassification decision into Z-scores (as standardized using the statewide mean and standard deviation of scale scores in each grade and year). This puts the scores onto a common scale across grades, which also makes coefficient sizes quite interpretable. California administered the CST to all students (except certain students in special education and some recently arrived immigrant students) in grades 2 through 11 from 2001-02 through 2012-13.

We also use CST Math scores for grades 3 through 7. At grade 8 and later, students take different CST math tests depending on their courses, and test scores are not comparable. For example, 9th graders may be taking either algebra or geometry.

Those EL students who narrowly miss reclassification are likely to be reclassified in the next one to three years, during which time they will continue to receive EL support services. The most obvious time-frame for measuring the ELA outcome is ELA performance in the first full school year after reclassification. As in a randomized controlled trial, we can gain precision by controlling for baseline achievement. As mentioned earlier, most students are reclassified part way through a school year (most typically in spring in SDUSD and throughout the year in LAUSD), and thus they receive the treatment for much but not all of the reclassification year. Were therefore set the baseline year to be the year before the reclassification decision.

On-time graduation can be measured based on the expected year of graduation, based on the grade in which the student is enrolled when his or her records are included in the RD sample for that year.

Methods

We use a regression discontinuity design (RD) that exploits the rules both districts have established for reclassification, as presented in Table 1. The assignment variables related to reclassification are the CST test scores and the overall and subtest scores on the CELDT and (in LAUSD) course marks. In both districts, a key distinction between what we call Era 1 and Era 2 is that in the latter era, a new more rigorous CELDT test was introduced. The other key distinctions are as follows. In San Diego, in elementary schools the CST cutoff was raised relative to Era 1. In Los Angeles, in Era 2 the district eliminated the requirement of reaching a certain grade in math classes. In San Diego, then, in Era 2 the reclassification unambiguously became more rigorous, due to the new CELDT test and, in elementary grades, a higher cutpoint on the state reading test (the CST). In Los Angeles, in Era 2 the two main changes worked in opposite directions, as the district dropped the math grade requirement at the same time that the state introduced a more demanding CELDT test. Thus it remains somewhat ambiguous in Los Angeles whether the overall standards became more or less rigorous.

Two questions must be addressed: What are the treatment and control groups, and is there a meaningful contrast between the experiences of the two groups? Students just below the cutoff point remain in EL support (described earlier), and are our control group. Students at or above the cutoff point are generally reclassified. The act of reclassification is meaningful. By

definition, reclassified students will be treated like otherwise similar native English speakers. Thus, the treatment is being reclassified and having English language development classes removed.

Conceptually, there are (at least) two ways to handle RD designs with more than one forcing variable, by combining the various criteria or by studying them separately. The What Works Clearinghouse (U.S. Department of Education, 2014) recommends that RDs for a given outcome, but based on different forcing variables, should be treated separately. We adopt this approach in our main analysis, which allows us to test whether cutpoints are appropriate separately for the various criteria.

Table 1 Reclassification Criteria in the Two Districts, and the Definition Of Eras in the Two Districts

	Los Angeles		San Diego
2003-06 (Era 1)	Basic ELA skills = 300+ CST ELA English proficiency (CELDT) = "Early Advanced" overall English proficiency = "Intermediate"+ subtests	**2004-05 (Era 1)**	Basic ELA skills = 300+ CST ELA English proficiency (CELDT) = "Early Advanced" overall English proficiency = no more than one "Intermediate" on subtests
	Teacher evaluation = 3+ or C or better in English *and* Math courses	**2005-06**	Increased CST ELA threshold to 333+

<div align="center">New CELDT era starting in 2006-07 school year</div>

2006-12 (Era 2)	Dropped math course requirement	**2007-14 (Era 2)**	333+ CST ELA for elementary, 300+ grades 6-12
2012-13	Allowed English course *or* Advanced ESL		
2013-14	Used 2012-13 CST *or* High school exit exam (grades 9-12) *or* Elementary reading assessment		

SOURCES: Los Angeles Unified School District. Reclassification of English Learners, Grades 2-12, Bulletin BUL-5619.0, October 17, 2011. San Diego Unified School District, Master Plan for English Learners, 2009

NOTES: When the new CELDT was introduced in 2006-07, cut scores for proficiency levels were raised across all grades and subtests (CDE, 2007). The two eras in each district are highlighted using lighter and darker shading for Eras 1 and 2 respectively. The range of years listed here and in later figures and tables refers to spring of the given school year. For example, the reference to "2003-2006" means that the school years 2002-2003 through 2005-2006 are used.

Estimation Method

A simple intent-to-treat estimator tests for whether there is a discontinuity in the outcome at the cutoff value of the running variable. Specifically, for the subsample of students who were ever English Learners, let t denote the year in which the reclassification decision is made, let $Y_{is,t+x}$ denote the test score or other outcome of student i in school s in school year t+x, which is x years after the reclassification decision where $x \geq 1$. Let $Z_{i,t-1}$ be a vector of baseline characteristics represent a vector of background variables measured the year before the reclassification decision. Note that the vector Z is measured in period t-1 because in the year of reclassification

reclassified students receive treatment for part of the school year before being reclassified. This vector includes time invariant demographic variables such as indicators for race/ethnicity, language spoken at home, and gender. But $Z_{i,t-1}$ also includes baseline student achievement characteristics. In terms of statistical expectation, there should be no differences in baseline characteristics of those just above and below the cutoff of the forcing variable, although in finite samples differences will emerge. By controlling for these variables we increase precision.

Consider the RD analysis based on the cutoff score of the ELA CST. The spring CST score is used to make a reclassification decision in the following school year. Thus, to be reclassified in year t, a necessary but not sufficient condition is that $CST_{i,t-1} \geq 0$ where we have rescaled the test score to equal 0 at the level required for reclassification in the given grade. In addition, the student must meet the other cutoffs imposed by the given district on the CELDT (and on course grades in LAUSD). The key regressor is a dummy variable $ABOVE_{it} = 1(CST_{i,t-1} \geq 0)$, (thus equaling 1/0 as the CST score is non-negative/negative). We estimate linear models on either side of the cutoff or, equivalently, estimate the two models at the same time by interacting controls with the ABOVE dummy:

$$Y_{is,t+x} = \alpha + \beta CST_{i,t-1} + \delta ABOVE_{it} + \gamma ABOVE_{it} \cdot CST_{i,t-1}$$
$$+ Z_{i,t-1}'\Delta + \mu_{i,t+x} \qquad (1)$$

Here, the key coefficient is δ. If it is not significantly different from zero, then we retain the null hypothesis of a zero causal impact of meeting the reclassification criterion on the outcome. (In the above model we assume a linear relation between the outcome and the running variable, $CST_{i,t-1}$, while allowing for different slopes on either side of the cutoff. In the main models we assume a more flexible quadratic model, and for robustness we later use higher order polynomials in the running variable as well. We focus on the quadratic and linear models in the main text as these models are less likely to overfit.)

The intent-to-treat model in (1) estimates the causal effect of meeting the reclassification criterion, but does not tell us the impact of treatment on the treated, that is, the impact of actual reclassification. Because we will have a fuzzy regression discontinuity design, based on Hahn et al. (2001) we can estimate the causal effect of reclassification, using a Two Stage Least Squares (2SLS) strategy. This approach produces a causal estimate of the impact of reclassification, which can be interpreted as a Local Average Treatment Effect (LATE). The coefficient of interest is a consistent estimate of the average causal effect of reclassification for ELs who were close to the cutpoint and who would comply with the reclassification policy.[12]

Our instrument for reclassification is the dummy variable $ABOVE_{it}$. To perform 2SLS, in the first stage we model the actual reclassification decision as

$$R_{it} = a + b\, ABOVE_{it} + Z_{i,t-1}'\Gamma + \epsilon_{it} \qquad (2)$$

where Γ is a vector of coefficients, a and b are coefficients, and ε_{it} is an error term. In the second stage, we model ELA achievement the year after the reclassification decision, but replace actual reclassification with predicted reclassification $\hat{R}_{it}$. We estimate linear models on either side of the cutoff or, equivalently, estimate the two models at the same time by interacting the running variable with the ABOVE dummy:

$$Y_{is,t+x} = \theta + \mu CST_{i,t-1} + \pi \hat{R}_{it} + \rho ABOVE_{it} \cdot CST_{i,t-1} + Z_{i,t-1}'\Lambda + \xi_{i,t+x} \qquad (3)$$

---

[12] Despite having a fuzzy RD because not all students are reclassified when they should be, we obtain consistent estimates under certain conditions (Angrist et al., 1996). The fact that not all students comply does not raise concerns of bias, but it does raise concerns about external validity (applicability to non-compliers), an issue that also exists in randomized controlled trials.

Here, the key coefficient is π. If it is not significantly different from zero, then we retain the null hypothesis of a zero causal impact of reclassification on test scores.

Tests for Manipulation of the Running Variable(s)

To have institutional integrity, the running variables such as the CST ELA score cannot be easily manipulated by teachers or other school officials who may take a personal interest in either reclassifying or not reclassifying a given student. Both the CST tests and the CELDT test are statewide tests, and they are graded outside of the given school district, which greatly reduces but does not eliminate the possibility of the scores being manipulated locally.

However, it is also useful to check for discontinuities in the density of the running variable at the cutoff point (McCrary, 2008). It is not required that the distribution be continuous at the cutoff to have a valid RD design (Imbens and Lemieux, 2008), but it increases confidence that no manipulation occurred.

A related check for manipulation of the running variable involves testing for a discontinuity in one or more baseline characteristics at the cutoff value of the running variable in the year before the reclassification decision is made.

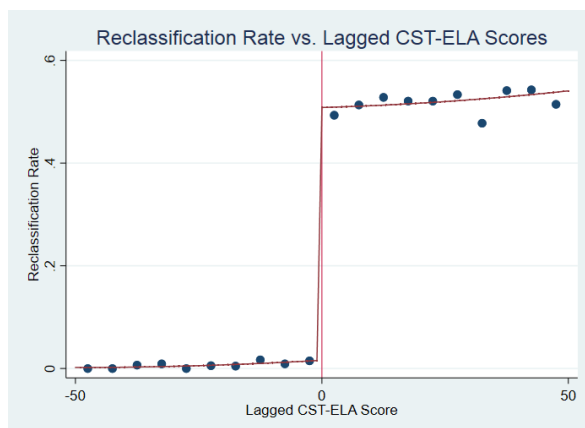# Results

Results of RD Validity Checks
We start by establishing whether the regression discontinuity design applies for the running variable (ELA CST scores), by checking for a discontinuity in the probability that a student is reclassified at the cutpoint. Figure 2 gives two graphical examples of the jump in reclassification rates from LAUSD and SDUSD, for elementary schools in the first era in each district. It shows the share of students reclassified versus the running variable, with the CST score rescaled to equal to zero at the cutpoint for the relevant grade. We see a large discontinuous jump of about 50 to 60 percentage points in the probability of being reclassified. Appendix Tables 1A and 1B show the results from the underlying models of the reclassification probability for the CST RD experiment, for samples used to model each outcome and for the elementary, middle and high school samples. In all cases we see a large discontinuity with the probability of being reclassified jumping by roughly 0.5 to 0.9. The lines in the graph show the regression fit using a quadratic polynomial estimated independently on the two sides of the cutoff. [13]

---

[13] In the main text we show results when fitting a second-order polynomial in the running variable. Results were similar whether using polynomials of degrees 1 through 4.   The samples used for these figures were the samples used to model the causal impact of reclassification on CST ELA scores in the year after the reclassification decision. Appendix Tables 1A and 1B show the coefficients on the discontinuity for this sample and for the other key outcomes: ELA scores two years later, math CST scores one and two years later, and graduation on time. Results are highly similar and always significant.

FIGURE 2 Probability of EL Reclassification at the CST Cutpoint, Grades 3-5

LAUSD Era 1 (2003-2006)   SDUSD Era 1 (2004-2005)

SOURCE: Authors' estimates

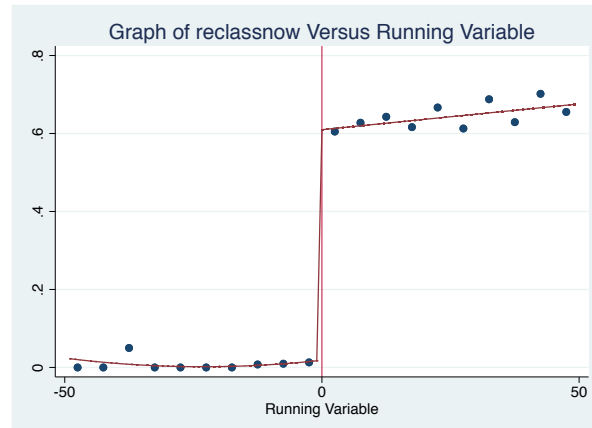NOTE: The lines in the graphs show the regression fit using a quadratic polynomial estimated independently on the two sides of the cutoff. The dots represent bin sizes of 5. The range of years listed here and in later figures and tables refers to spring of the given school year. For example, the graph for LAUSD for "2003-2006" means that the school years 2002-2003 through 2005-2006 are used.

We next turn to checks on the validity of the RD design. We begin by performing the McCrary test for manipulation of the running variable. We expect to see no discontinuity in the frequency of students above versus below the cutpoint. Tables 2A and 2B in the appendix show results for this test for the sample used to model the impact of reclassification on CST ELA scores the year after reclassification. (We use counts of students in intervals of 5 points on the CST ELA test.) The tables also show the same for the sample used to model the impact of graduation on time. [14] The coefficient on ABOVE is not significantly different from zero in either subsample, for any of the three grade spans. Figures 3A and 3B give one example from each district, showing the distribution of students versus the running variable for elementary grades in the more recent ("new") eras in either district.

---

[14] This second set of checks using graduation on time is really a test not just for manipulation of the CST test scores, but of differential attrition in the one to nine years between the reclassification decision and when students are expected to be in grade 12.

Figure 3 Frequency of Observations at the Cutoff, for Era 2, CST Running Variable
Figure 3A LAUSD 2007-12 Elementary grades (3-5)
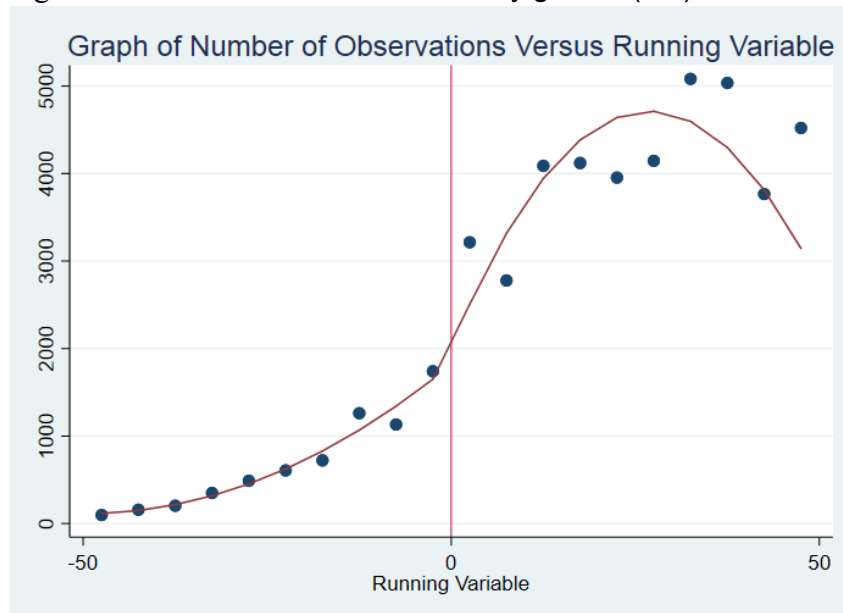
Figure 3B SDUSD, 2007-2014 Elementary grades (3-5)
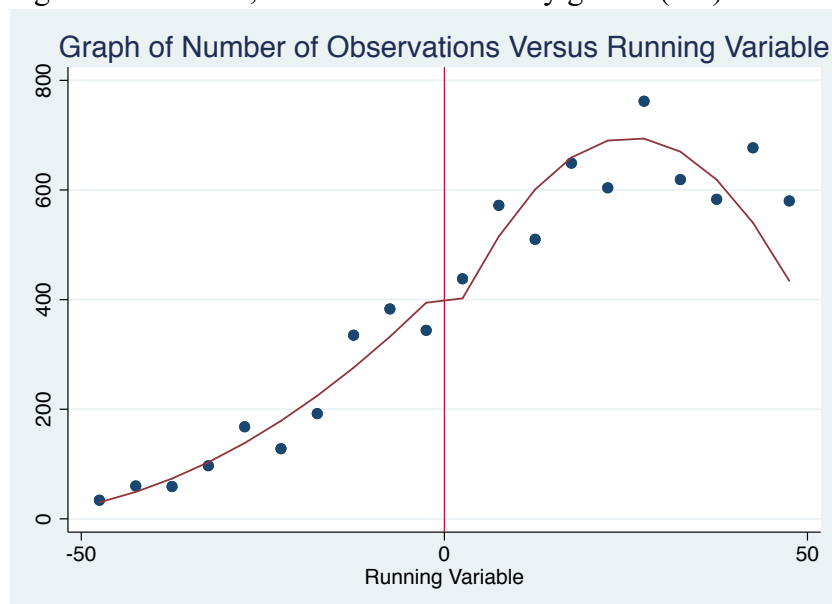
Notes: The samples for both districts are students in the Grade 3-5 RD sample for the model of the impact of meeting the CST cutoff.   The dots represent bin sizes of 5.

While not strictly required for the RD design to be valid, a finding that there is not any discontinuous jump in the mean value of each background variable at the CST cutoff would provide reassurance that the treatment and control groups are similar. Tables 3A and 3B in the appendix show the coefficient on ABOVE in models where the dependent variable is one of a number of student background characteristics.

For SDUSD, the discontinuity variable is almost always statistically insignificant, with a few exceptions.

For LAUSD, slightly more of these tests produce significant discontinuities than one would expect by chance. In LAUSD, there are no discontinuities observed in the numbers of students who are female or Spanish speakers just above the CST cutoff. However, there are a number of negative discontinuities observed just above the CST cutoff for the various CELDT subtests (especially the reading subtests), as well as with the overall score on the CELDT. The differences are modest, when expressed as proportions of a standard deviation. (Standard deviations appear in Appendix Table 4A). For example, in middle school in Era 2, those above the cutoff had a CELDT overall score in the prior year 5.5 points lower than those just below the cutoff, which translates to a difference of 0.09 standard deviations. Academic year and grade level are often statistically significant as well, as they sometimes are in SDUSD.

Overall, it appears that both districts have adhered to their stated reclassification policies quite closely, and that the fuzzy RD design is appropriate. That said, it seems sensible to control for baseline characteristics to increase precision, especially in LAUSD.

CST Experiments

Turning now to the estimated impacts, Tables 2A and 2B show the estimates for the intent to treat effect. Appendix Tables 4A and 4B show sample sizes and means and standard deviations of the various outcomes and explanatory variables. Appendix Tables 5A and 5B show these models as well the effect of treatment on the treated. Patterns of statistical significance and signs are identical between the intent-to-treat and the estimates of the impact of treatment on the treated, with the latter mechanically equaling the former divided by the difference in the share of students who are reclassified for those above and below the cutoff.

Tables 2A and 2B show the coefficient and standard error on ABOVE for the quadratic model. [15] We interpret an insignificant coefficient as meaning that students are being reclassified at the right time – although EL supports are removed for the reclassified student he or she performs as well as those who receive an extra year of EL supports. But a (negative/positive) and significant coefficient would imply that the student was being reclassified too (soon/late). The tables show five outcomes – CST ELA and math test scores for one and two years after the reclassification decision, and a longer-term outcome – whether the student graduated on time.

---

[15] (Appendix Tables 5A and 5B also show the results when the running variable is included on the right-hand side as polynomials of degrees 1 (linear) through 4. We discuss these results in more detail in the Robustness section below.)

Table 2A Quadratic Polynomial Results for LAUSD CST-ELA Experiment (Intent to Treat)

		CELDT Era 1			CELDT Era 2		
		Elementary	Middle	High	Elementary	Middle	High
CST ELA 1 year ahead	Coef.	-0.0286	0.0085	0.0033	**-0.0742****	0.0112	0.0145
	S.E.	(0.0162)	(0.0134)	(0.0196)	(0.0172)	(0.0124)	(0.0194)
	# Obs	40,141	37,577	16,769	48,361	40,361	20,470
CST ELA 2 years ahead	Coef.	**-0.0357***	-0.0271	-0.0209	0.0025	0.0013	-0.0245
	S.E.	(0.0177)	(0.0149)	(0.0291)	(0.0202)	(0.0153)	(0.0303)
	# Obs	36,927	34,464	8,799	38,227	32,058	9,269
CST math 1 year ahead	Coef.	0.0105	0.0243		-0.0247	0.0077	
	S.E.	(0.0194)	(0.0298)		(0.0203)	(0.0291)	
	# Obs	40,121	13,788		48,351	13,043	
CST math 2 years ahead	Coef.	0.0237			-0.0027		
	S.E.	(0.0201)			(0.0227)		
	# Obs	36,773			38,184		
Graduation on time	Coef.	**-0.0315***	0.0148	**-0.0319***	-0.0044	-0.0035	0.0090
	S.E.	(0.0157)	(0.013)	(0.0136)	(0.0219)	(0.0136)	(0.0119)
	# Obs	23,444	24,809	22,979	13,179	23,795	31,079

SOURCES: Authors' calculations.

NOTES: For each outcome the table shows the coefficient that measures the size of the discontinuity, the standard error and the number of observations in the model. The table shows intent-to-treat estimates. Impacts of treatment on the treated are similar but larger in proportion to the increase in the share of students reclassified above the cutpoint. Model of CST Math 1 Year Later for middle school students is only for students whose reclassification decision was in grade 6. The models shown here use a quadratic polynomial estimated separately on either side of the discontinuity. **: Significant at the 1% level; *: Significant at the 5% level.

Table 2B Quadratic Polynomial Results for SDUSD CST-ELA Experiment (Intent to Treat)

		CELDT Era 1			CELDT Era 2		
		Elementary	Middle	High	Elementary	Middle	High
CST ELA 1 year ahead	Coef.	-0.0034	-0.033	**-0.138***	-0.0087	-0.0185	-0.0557
	S.E.	(0.0526)	(0.0418)	(0.0668)	(0.0362)	(0.0400)	(0.0663)
	# Obs	3,414	3,903	1,513	7,926	6,540	1,931
CST ELA 2 years ahead	Coef.	**-0.124***	-0.0649	0.129	-0.0453	**0.114***	-0.0515
	S.E.	(0.0582)	(0.0484)	(0.0877)	(0.0435)	(0.0466)	(0.118)
	# Obs	3,182	3,607	863	5,954	5,155	746
CST math 1 year ahead	Coef.	0.0302	-0.0267		-0.0712	-0.0709	
	S.E.	(0.0644)	(0.110)		(0.0474)	(0.106)	
	# Obs	3,414	1,054		7,922	2,044	
CST math 2 years ahead	Coef.	-0.0651			-0.0460		
	S.E.	(0.0660)			(0.0570)		
	# Obs	3,172			5,781		
Graduation on time	Coef.	0.0666	-0.0398	0.0657	0.0289	-0.0098	0.0096
	S.E.	(0.0582)	(0.0454)	(0.0605)	(0.0550)	(0.0393)	(0.0309)
	# Obs	2,702	3,527	1,636	1,814	4,271	5,715

SOURCES: Authors' calculations.

NOTES: For each outcome the table shows the coefficient that measures the size of the discontinuity, the standard error and the number of observations in the model. The table shows intent-to-treat estimates. Impacts of treatment on the treated are similar but larger in proportion to the increase in the share of students reclassified above the cutpoint. Model of CST Math 1 Year Later for middle school students is only for students whose reclassification decision was in grade 6. The models shown here use a quadratic polynomial estimated separately on either side of the discontinuity. **: Significant at the 1% level; *: Significant at the 5% level.

We find that in Los Angeles Unified, there are a few instances when it appears that students were reclassified too soon.[16] High school students reclassified in the 2003-06 era saw negative impacts on their chances of on time graduation. Those at or above the reclassification cutoff had approximately 3% lower chances than those students below the reclassification cutoffs. As shown in Appendix Table 5A, the impact of treatment on the treated, that is, the causal effect of being reclassified in high school, is minus five percentage points. The results suggest that there were longer-term consequences of being reclassified too soon in LAUSD. High school students who were just above the CST cutpoint had slightly lower graduation rates than those students who were just below the graduation cutpoint.[17] Given that many of the high school students who narrowly missed the cutoff were reclassified a year or two later, this estimated drop in graduation

---

[16] As stated above, we estimated linear and quadratic models, 3rd and 4th order polynomial models. Our estimates on being above the reclassification threshold were usually but not always consistent. See appendix tables 5A and 5B for the full results. The most common pattern where the four models differed in significance was that the simple linear model sometimes disagreed with the models with higher order polynomials. We interpret this as most likely being due to some non-linearity in the relation between the running variable and the given outcome, which is not allowed for in the linear model.

[17] It is important to note that in LAUSD, if a student exits the school district, they are absent from our sample, whereas in SDUSD, the district tracks students who are dropouts. Therefore, LAUSD's on-time graduation outcome in this study is calculated among students who persist to 12th grade. On-time graduation in SDUSD includes dropouts in the denominator.

rates is quite significant from a policy viewpoint. The result implies that one or two additional years of EL supports for students close to the cutoff could reduce their timely progress on the courses needed for high school graduation. Robinson-Cimpian and Thompson (2016) report similar results for graduation in LAUSD.[18] However, we note that this drop in graduation rates on time occurs when we used models with polynomials of orders 2, 3 and 4, in the simplest linear model the coefficient, while still negative, is closer to zero and is no longer statistically significant.

In "Era 2" in LAUSD, after the CELDT was renormed and the math course grade requirement was dropped, there were slightly different measurable impacts of reclassification in LAUSD. The negative high school graduation effect disappears, suggesting that perhaps the more rigorous CELDT requirement was more important than removing the math course requirement in improving reclassification timing with respect to the high school graduation outcome. We also find that elementary school students who were reclassified in the 2007-12 era had lower CST ELA scores one year after reclassification.[19] However, two years after reclassification, reclassified students appear quite similar to those who just missed being reclassified two years earlier. Conversely, in Era 1 elementary school students who met the reclassification cutoff had similar scores to those who did not meet the cutoff one year later, but by two years later the effect became negative and significant.

In SDUSD, in most cases those who qualified for reclassification by meeting the CST cutoff performed the same as those who were just below the cutoff. This suggests SDUSD set its reclassification criteria appropriately. There were two exceptions. In Era 1, when the first version of the CELDT test was in place in 2003-04 and 2004-05, those meeting the CST cutoff in high school had lower CST ELA scores a year later. In Era 2, where we study reclassification decisions made in 2006-07 through 2013-14, middle school students meeting the CST cutoff performed better on the CST ELA test two years after the reclassification decision than those just below the cutoff. One should not make too much of two exceptions out of 26 tests for appropriate cutoffs. Nonetheless, it is notable that the one case indicating students might be reclassified too soon came from Era 1 when the less rigorous CELDT was in place, and the one case of possible reclassification too late came in Era 2 when the more rigorous CELDT test was in place.

Figures 4 through 6 show the discontinuities for some of the cases in which we find that students perform worse when reclassified than those just below the cutpoint who are not reclassified. Figures 4 and 6 quite clearly show a change in the outcome at the discontinuity, the first in LAUSD and the second in SDUSD. Figure 5 shows the discontinuity in the rate of on-time graduation in LAUSD. As mentioned earlier, this effect is negative and significant when we use polynomials of order 2, 3 and 4 in the running variable, but in a linear model the effect, while still negative is not significant. The graph shows that the discontinuity is not as clear-cut as in the other cases shown.

---

[18] Robinson-Cimpian and Thompson (2016) find a negative impact of 11 percentage points on graduation for high schoolers reclassified in 2003-2006.
[19] The math results were not statistically significant for the second-order polynomial, but were for the other models.

Figure 4 Regression Discontinuity in CST ELA Scores One Year after Reclassification Decision in LAUSD, Grades 3-5, 2007-12
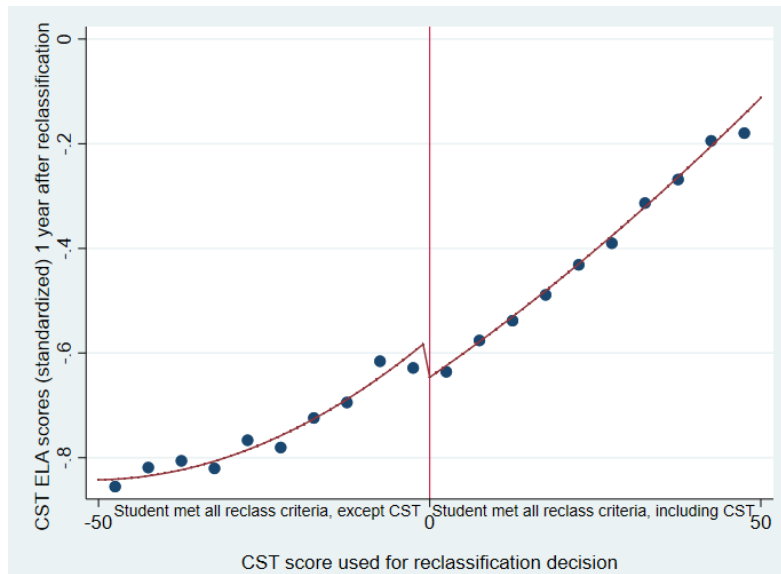
Figure 5 Regression Discontinuity in the Probability of Graduation on Time after Reclassification Decision in LAUSD, grades 9-12, 2003-06
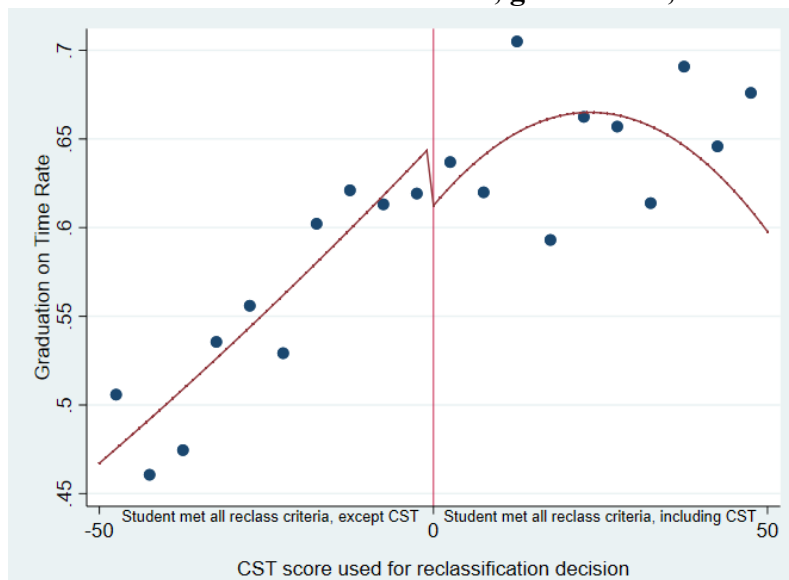
Figure 6 Regression Discontinuity for CST ELA Scores One Year After Reclassification Decision in SDUSD, Grades 9-12, 2004-2005

Robustness

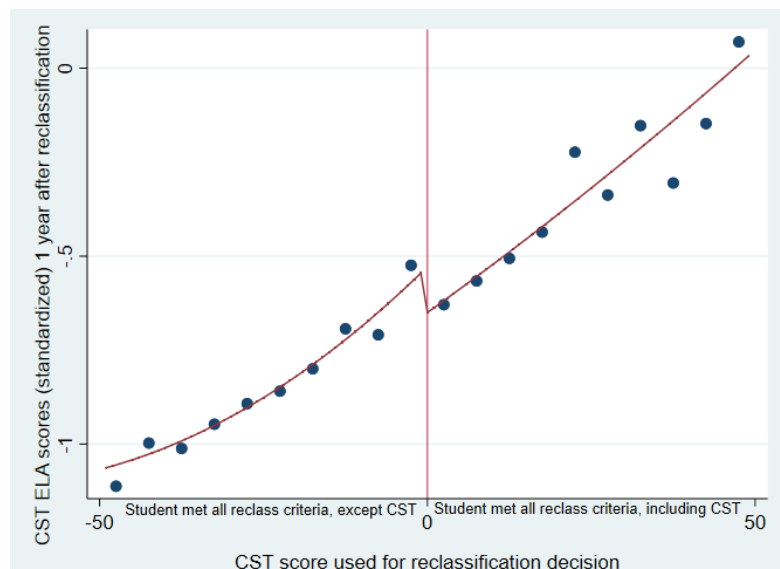One potential issue is the order of the polynomial of the running variable. Visual inspection of the graphs and levels of significance of the various terms in the polynomial persuaded us to use the second order polynomial for the main text because of some apparent slight non-linearities in the graphs, but we also used linear, third- and fourth-order polynomials. A concern with the latter two polynomials is that we could inadvertently overfit the data or give undue weight to a few observations far from the cutpoint, but we include them for completeness. Appendix Tables 5A and 5B replicate the intent-to-treat and treatment-on-the-treated results for each outcome, and using each of these polynomials. In many cases, the results are insensitive to the choice of the order of the polynomial. We have already mentioned one important exception, where the negative impact of reclassification on the probability of on-time graduation in LAUSD in Era 1 is significant in all three non-linear specifications, but not in the linear specification. There were seven other cases in LAUSD where the estimated impact differed between the linear and quadratic models. In all but one of these cases the estimated impact was negative in one case and statistically insignificant in the other case. Furthermore, six of these seven cases in LAUSD occurred in Era 1. Although our interpretation is that we must take these impacts as uncertain, they do point to additional evidence that in Era 1 in LAUSD students were in some cases reclassified too early.

Another potential issue is whether the bandwidth influences the findings. Wider bandwidths will generate greater precision but potentially at the cost of greater bias. In results available from the authors we reproduced the results here, which used a bandwidth of +-50 scale score points, using bandwidths of +-100 and +-30 points. Results are broadly similar, although in a few cases with the smallest bandwidth results become statistically insignificant, even though the sign of the

estimated impact does not change. We believe that this merely represents a loss of precision as we trim the sample.

CELDT Experiments

In LAUSD, we were able to conduct similar experiments in which we considered the impact of the requirement that students score at an overall Early Advanced performance level on the CELDT Overall Proficiency Level.[20] We performed similar validity checks to our CST experiments and concluded that, in most cases, we could confidently measure an effect of being just above the cutpoint.[21] In these experiments, we considered students who met the CST requirement, earned a course mark of "C" or better in English and math in 2003–06, and met all CELDT subtest requirements by scoring at least intermediate on reading, writing, listening, and speaking. Correlations among the CELDT subtests and the overall CELDT score are high, but 27 percent of students meeting all CELDT subtest criteria did not meet the overall CELDT cutpoint by having an overall score of Early Advanced. We then compared students just above and just below the CELDT Early Advanced cutpoint.[22] It appears that the district reclassified students according to policy at very high levels in Era 2. Those just meeting the CELDT overall requirement had between 61 and 85 percentage point gains in the reclassification rate in the new era. Adherence to policy was weaker in Era 1, with reclassification rate jumps between 20 and 50 percentage points.[23,24]

In Era 1, there were no significant effects of meeting the CELDT OPL cutoff at any grade level on CST ELA scores one or two years after reclassification, on CST math scores one or two years after reclassification, or on on-time graduation (see Table 3). This implies that the CELDT Early Advanced overall performance requirement was set approximately correctly for smooth transition from EL to reclassified status given the other reclassification requirements.

Once the CELDT became more challenging in 2006–07 and LAUSD dropped the math grade requirements, reclassified middle school students who scored just above the CELDT OPL had slightly lower CST ELA scores than similar students who scored just below the cutpoint in the one, but not two years later. Given both the short time frame of this finding and that some of our tests failed to confirm RD validity for middle school students reclassified in Era 2, we do not put too much stock in this finding.

---

[20] Cutpoints for the Early Advanced level are pegged to scaled scores. Scaled scores required for Early Advanced vary by grade level.

[21] There were a few cases in which these tests suggested some imbalances in student characteristics or numbers of students at the discontinuity. Middle school students just above the CELDT OPL cutpoint in Era 2 had statistically higher CELDT subtests scores than students just below the cutpoint. Similarly, the McCrary test suggested that the RD experiment may not be valid for students reclassified in middle school in Era 2.

[22] Validity checks are available from the authors on request.

[23] However, we do not have confidence in the results for middle school reclassification in Era 1 when it comes to our estimates of the impact on math scores. The size of the discontinuity in the probability of reclassification at the CELTD OPL cutpoint for the math outcomes was just 0.20.

[24] Reclassification decisions in LAUSD are now largely automated.

Table 3 Quadratic Polynomial Results for LAUSD CELDT-OPL Experiment (Intent to Treat)

		CELDT Era 1			CELDT Era 2		
		Elementary	Middle	High	Elementary	Middle	High
CST ELA 1 year ahead	Coef.	-0.0048	0.0020	0.0440	-0.0201	**-0.0410***	-0.0151
	S.E.	(0.0165)	(0.0282)	(0.0580)	(0.0116)	(0.0195)	(0.0324)
	# Obs	39,807	16,566	5,307	71,172	21,689	9,724
CST ELA 2 years ahead	Coef.	-0.0052	0.0284	0.0782	-0.0236	-0.0359	0.0005
	S.E.	(0.0189)	(0.0291)	(0.0752)	(0.0139)	(0.0230)	(0.0527)
	# Obs	36,843	15,401	2,700	55,122	17,087	3,656
CST math 1 year ahead	Coef.	-0.0124	0.0079		0.0057	-0.0246	
	S.E.	(0.0222)	(0.0409)		(0.0154)	(0.0353)	
	# Obs	39,706	9,543		71,133	10,359	
CST math 2 years ahead	Coef.	-0.0040			0.0123		
	S.E.	(0.0230)			(0.0177)		
	# Obs	36,755			55,060		
Graduation on time	Coef.	(0.0210)	(0.0245)	(0.0463)	0.0109	0.0075	0.0143
	S.E.	-0.0147	-0.0219	-0.0374	(0.0170)	(0.0202)	(0.0193)
	# Obs	23,219	11,049	6,736	14,333	11,786	13,407

SOURCES: Authors' calculations.

NOTES: For each outcome the table shows the coefficient that measures the size of the discontinuity, the standard error and the number of observations in the model. The table shows intent-to-treat estimates. Impacts of treatment on the treated are similar but larger in proportion to the increase in the share of students reclassified above the cutpoint. Model of CST Math 1 Year Later for middle school students is only for students whose reclassification decision was in grade 6. The models shown here use a quadratic polynomial estimated separately on either side of the discontinuity. **: Significant at the 1% level; *: Significant at the 5% level.

In SDUSD, with only a handful of students who met all other criteria but were just below the OPL cutpoint, there were not enough observations to draw meaningful conclusions. This is understandable because SDUSD required students to be Early Advanced on all but one CELDT subtests, with an Intermediate on at most one subtest. Mechanically, this reduced the possibility of being at these levels on the subtests but below Early Advanced on the Overall Proficiency Level.

Both districts had enough students above and below the CELDT reading subtest cutpoint to permit examination of the effect of reclassification based on meeting the CELDT reading

criterion in at least some grade levels.[25] For the CELDT reading subtest, as with the CELDT OPL, LAUSD reclassification policy seems to have been implemented more regularly in Era 2. In this era, the probability of reclassification jumps at or above the CELDT reading cutpoint 68 to 87 percentage points compared with 31 to 47 in Era 1.

While the CELDT OPL cutpoint is Early Advanced in LAUSD, scores on any subtests need only be Intermediate for an EL student to be eligible for reclassification. We had enough students on either side of the cutpoint to estimate the effect of being just above the cutpoint for elementary students in the first reclassification era and elementary, middle, and high school students in the second era. Table 4A, shows the results. We found no impact on our academic outcomes of interest and conclude that, given the other reclassification requirements, CELDT reading cutpoints for elementary school students were set at the proper level in both reclassification eras. There was one case with a significant discontinuity in the new era in LAUSD, with a positive jump in the probability of graduating on time for those meeting the CELDT Reading cutoff in middle school. However this effect was significant only in the quadratic polynomial and not for the linear or the higher order polynomials. We conclude that overall for middle and high school students[26], CELDT reading cutpoints were also properly set in the new era.

Table 4A Quadratic Polynomial Results for LAUSD CELDT-Read Experiment (Intent to Treat)

		CELDT Era 1			CELDT Era 2		
		Elementary	Middle	High	Elementary	Middle	High
CST ELA 1 year ahead	Coef.	0.00455			0.0214	-0.0453	0.00479
	S.E.	(0.0280)			(0.0289)	(0.0511)	(0.0559)
	# Obs	22,157			45,735	9,007	5,016
CST ELA 2 years ahead	Coef.	0.0481			0.0503	-0.113	0.0346
	S.E.	(0.0306)			(0.0359)	(0.0678)	(0.0846)
	# Obs	20,578			35,508	7,226	1,798
CST math 1 year ahead	Coef.	-0.0316			0.0159	-0.0445	
	S.E.	(0.0377)			(0.0395)	(0.0936)	
	# Obs	22,150			45,720	4,365	
CST math 2 years ahead	Coef.	-0.00559			-0.0300		
	S.E.	(0.0391)			(0.0466)		
	# Obs	20,531			35,472		

---

[25] Validity checks are available from the authors on request. The McCrary for the CELDT reading test in LAUSD showed no bunching of observations on either side of the cutoff. In LAUSD, there are occasional discontinuities observed in the numbers of students who are female or Spanish speakers just above the CELDT reading cutoff, but they are not consistent across outcomes and grade levels, nor across specification. We control for these background characteristics in our RD experiments. For the SDUSD CELDT reading experiment in Era 2, the McCrary tests showed no evidence of bunching. Only a few of the baseline characteristics showed a discontinuity, and the discontinuities each occurred in only one of three grade-spans. The most notable imbalance was that in high school and for the sample used to model CST ELA scores one year later, those above the cutoff were 16% more likely to have Spanish as their home language than those just below the cutpoint. But no such gap emerged in the lower grades. As another example, the mean year of the reclassification decision was lower by 0.4 in middle school for those just above the cutpoint, but there was no difference for the other grade-spans. As a precaution, and in line with the other analyses in the paper, we control for all of baseline variables in the RD outcome models.

[26] High school graduation rates were higher for those just above the reclassification cutpoint in high school, but only in the quadratic model, not in the linear or other polynomial models.

Graduation on time	Coef.	-0.00120			-0.0673	**0.135***	-0.000462
	S.E.	(0.0261)			(0.0631)	(0.0612)	(0.0332)
	# Obs	12,743			8,796	4,949	6,601

Table 4B shows results of the CELDT reading analysis in SDUSD during Era 2. In all cases we find no evidence that students were being reclassified at the wrong time. However, the sample size is rather small for the high school model of outcomes two years ahead on the CST reading test. In that case we have less confidence that the evidence clearly shows that the cutpoint was appropriate.

Table 4B Quadratic Polynomial Results for SDUSD CELDT-Read Experiment (Intent to Treat)

		CELDT Era 2		
		Elementary	Middle	High
CST ELA 1 year ahead	Coef.	-0.0352	-0.0154	0.0429
	S.E.	(0.0231)	(0.0349)	(0.0679)
	# Obs	10,953	6,082	1,674
CST ELA 2 years ahead	Coef.	0.0417	-0.0198	-0.206
	S.E.	(0.0281)	(0.0410)	(0.118)
	# Obs	7,968	4,745	602
CST math 1 year ahead	Coef.	-0.0202	0.0822	
	S.E.	(0.0321)	(0.0697)	
	# Obs	10,950	2,494	
CST math 2 years ahead	Coef.	-0.00948		
	S.E.	(0.0393)		
	# Obs	7,776		
Graduation on time	Coef.	0.0557	-0.0167	-0.00220
	S.E.	(0.0404)	(0.0356)	(0.0421)
	# Obs	2,027	3,757	2,787

Does reclassification have different effects on long-term and late-arriving English Learners?

At the secondary school level, ELs are heterogeneous in the length of time they are in the district before reclassification. Long Term English Learners (LTELs) are those who have already completed at least five years of schooling and have not been reclassified.[27] Late Arriving English Learners are those arriving for the first time at grade 6 or higher while scoring at the bottom proficiency level on the CELDT (Beginning level) when they first take the test. Hill et al. (2019) study these two populations in LAUSD and SDUSD, and finds that late-arriving ELs have much lower reading achievement than LTELs, but on average make faster gains. In both districts, LTELs greatly outnumber late-arriving ELs. A third, and very small, group consists of ELs who arrive before grade 6 but have not yet spent five years as ELs in a school district by the end of grade 6.

In LAUSD, we examined reclassification effects among LTELs in grades 6-8 and 9-12. On the CST, we found that a negative effect of reclassification on high school graduation was only weakly evident in one of the four models we estimated. Otherwise, LTELs just above the CST cutpoint had similar academic outcomes. Our estimates for LTEL academic outcomes after reclassification using the CELDT overall and reading scores also failed to find any statistically significant differences.

Because the numbers of late arrivers are relatively small, we combined grades 6-12 to estimate as many regression discontinuity experiments as possible for this group in LAUSD. In all instances for which we had enough students to perform the experiment, we found no statistically significant differences for students just above and below the cutpoint.[28]

In SDUSD, we lacked a sufficiently large sample of late-arriving ELs to repeat the analyses for this group. But we did have enough LTELs to repeat all of the experiments for this subgroup in grades 6-8 and 9-12. The results were extraordinarily similar to the main results described above. In no case was the impact of being above the cutoff or of being reclassified statistically significant. Furthermore, in the few cases where there was a significant impact for all ELs, the direction of the estimated effect was the same. LTELs constituted well over half the samples used in our earlier analyses, which may explain why findings were similar when we used all ELs in middle and high schools.

We conclude that separate reclassification policies for late-arriving and long-term ELs do not appear to be necessary, though we caution that, even in LAUSD, our sample of late-arriving ELs is small.

---

[27] In SDUSD, this is defined as 5 years in the district and in LAUSD, it is defined as 5 years in the U.S.
[28] In the CST experiment, we were able to estimate CST ELA scores one year post reclassification in Era 2 and high school graduation in both eras. For the CEDLT OPL experiment, we were able to estimate CST ELA scores and gradation in Era 2. For the CELDT Reading experiment, we were only able to estimate high school graduation in Era 2.

Do moderating factors impact the effect of reclassification?

The appropriate timing of reclassification may vary according to student, school, or neighborhood characteristics. If so, the optimal reclassification strategy could vary depending on student subgroups or school contexts. We examine five types of potential factors that might affect our CST ELA test score experiment one year later:

1. language spoken at home;

2. student demographics, including concentrations of students by language and diversity of languages spoken at a school, the latter calculated as a Herfindahl index;

3. neighborhood characteristics, such as language, poverty, and nativity in the school catchment area, based on 2006–16 zip code data estimates from the American Community Survey;

4. average teacher qualifications at a school;[29]

5. fidelity to English Language Development course placement as outlined in each district's *Master Plan for English Learners*.[30]

In both districts, we estimated separate regression models with one moderator included at a time to determine if any of these moderating factors appears to have an impact on reclassification. For each moderator, we estimated a model that added the moderator plus an interaction between the moderator and the indicator for being at or above the cutpoint on the running variable.

LAUSD Results

In LAUSD, all the following variables appeared to have a negative impact on CST ELA scores for students just above the CST reclassification cutpoint relative to those just below, when estimated in separate regression models (see Appendix Table 6A):

- being a Spanish speaker, as compared to speakers of other languages;

- school language characteristics, including overall EL percentage, the percentage of ELs speaking the same language as the student, the percentage of the overall student body speaking the same language as the student, and a measure of language homogeneity among EL students;

- neighborhood language characteristics, including the percentages of foreign-born and Spanish-speaking people.

However, these findings must be considered tentative because the measures of language are closely correlated. We then estimated a model with just one variable from each of the five types of moderating factors listed above, only focusing on language characteristics, and included these potentially moderating factors simultaneously. We found little evidence of a systematic effect of

---

[29] We measure average teacher tenure and percentage of teachers with at least a master's degree at each school among all teachers, using publicly available data from the California Department of Education.
[30] Each school district has well-documented policies for the placement and advancement of EL students through EL coursework. Using this documentation as guidance, we calculated whether individual EL students were properly placed in EL instruction each year and what share of EL students at each school were properly placed each year. See Hill et al. (2019) for additional detail.

moderators on CST ELA test scores across reclassification eras or grade levels. (See Appendix Table 7.)

For example, in Era 1 in LAUSD, only two of the 15 interactions between the moderating variables and the indicator for being above the cutoff were statistically significant. We do not see enough evidence to draw conclusions about the effects of these variables in Era 1. In the more recent reclassification era, more of these variables seem to affect reclassification outcomes for students reclassified in middle and high school. However, these results are either not robust to variations in student grade level or reclassification era, or they are largely offset by reclassification's positive overall effect. Overall, these results do not appear to provide a compelling reason to modify reclassification policy according to school context or linguistic subgroup, but they do provide some indications that the effects of reaching the reclassification cutoff could be smaller for some subgroups.

SDUSD Results

In San Diego, moderating variables that measured students, schools, and neighborhood characteristics did not for the most part appear to have an independent effect on ELA standardized test scores one year after reclassification, an important outcome. Appendix Table 6B provides the results.

There were a few exceptions in the CST ELA experiment. We found a negative interaction with an indicator for Spanish as the home language for two of three grade spans in Era 2 and one grade span in Era 1. This is similar to our findings for LAUSD. Second, for both elementary and middle schools in Era 2, the effect of meeting the CST reclassification cutoff was smaller if teachers had more experience. One interpretation is that, in that era, it might have been worthwhile postponing reclassification for students in schools with highly experienced teachers.

There were three other moderating variables that were significant, but this pattern appeared in only one grade span and one era, which raises concerns that the results could have been accidental.[31]

We tentatively conclude there is little evidence that student, school, or neighborhood characteristics had a separate effect on academic outcomes as a result of reclassification. This probably is good news for educators and policymakers who must standardize EL reclassification criteria across the state. The exception for which both districts provided some evidence is that meeting the reclassification cutpoint had smaller effects for ELs whose home language was Spanish, relative to other ELs. In addition, there may other characteristics of the EL instructional environment or the reclassified student environment that we do not measure that could be relevant. Unmeasured differences across students could also contribute to an explanation about why reclassification criteria can differ across LAUSD and SDUSD but still be appropriate for students in their districts.

---

[31] First, in Era 2 for middle schools, the proportion of teachers holding a Master's degree had negative and significant interaction with the indicator for meeting the CST criterion. Second, in Era 1 there was a negative and significant interaction only between the indicator for whether the student's home language was Spanish and meeting the cutoff in high school. This was also found in the LAUSD results, but in that case the negative interaction was in Era 2. Third, in Era 2, the impact of meeting the CST cutoff was smaller for high school students as the share of foreign-born rose in the school zip code. Notably, this pattern was not statistically significant when we used the share of foreign-born in the zip codes where students lived.

Conclusion

We began this paper by noting that EL reclassification is one of many examples where a government agency provides supports to a subset of people who have particular needs, with the clear intention of providing these supports only temporarily. This paper provides an example of how RD designs can help policymakers determine whether their criteria remove the temporary supports at the optimal time. This approach could be used in many different settings.

Overall, in the two school districts and two reclassification eras we studied, reclassification criteria appear to largely have been appropriate. We found one reclassification policy, in LAUSD in Era 1, where EL supports appear to have been removed too soon. But when the state introduced a new more rigorous language test, this sufficed to solve the problem of reclassifying ELs too soon.

In SDUSD, the second largest district in California after LAUSD, almost without exception we found that those reclassified performed about as well those just below the cutpoint. This poses a puzzle because the state increased the rigor of its language proficiency test in Era 2 we studied for this district, and at the elementary level the district additionally increased its requirements in Era 2 beyond the change in the state test. Implicitly, we had imagined a single optimal set of criteria for reclassification would exist. These results, though, instead suggest that the district was operating on the "flat of the curve" for gains from reclassifying students. In other words, there is likely to be a *range* of reclassification criteria cutpoints that all represent appropriate times to remove EL support services from students.

Educators and policymakers have long understood that it makes no sense for similar students to have different EL status depending on where they attend school. Indeed, a survey of California districts by Hill, Weston, and Hayes (2014) revealed considerable variation in reclassification policies among districts. This idea of the need to create a common set of rules for reclassifying students has trickled up to the federal level. Standardization of reclassification policies is required by the federal Every Student Succeeds Act (ESSA), signed into law in 2015. ESSA is likely to create considerable disruption as districts in each state abandon their former policies in favor of new reclassification policies that individual states are now designing.

A benefit of the standardization required by ESSA, though, is that once reclassification policy is standardized across school districts within each state, we may learn more about the effects of different district instructional models.

For the many states with large EL populations it will be important to assess whether states' new standards, which are currently being designed and rolled out, are set appropriately. Monitoring recently reclassified students, as is required by ESSA, is one important way to do so, but regression discontinuity modeling represents an ideal approach to answering the question of whether the policies are the optimal ones.

References

Angrist, Joshua D., Imbens, Guido W., & Rubin, Donald B. (1996). "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*: 91(434), 444-455.

Callahan, Rebecca. 2005. "Tracking and High School English Learners: Limiting Opportunity to Learn." American Educational Research Journal 42 (2): 305-328.

California Education Code, Section 313(f).

California Department of Education (2019). Designated and Integrated English Language Development. Memo, retrieved September 24, 2019.

Carlson, Deven, and Jared E Knowles. 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 & Management* 35 (3).

Cimpian, Joseph R, Karen D Thompson, and Martha B Makowski. 2017. "Evaluating English Learner Reclassification Policy Effects Across Districts." *American Educational Research Journal* 54 (1): 255.

EdSource. 2008. *English Learners in California: What the Numbers Say.* Mountain View, CA.

Flores, E., G. Painter, and H. Pachon. 2009. *¿Qué Pasa?: Are English Language Learning Students Remaining in English Learning Classes Too Long?* Los Angeles: Tomás Rivera Policy Institute.

Gándara, P., and R. Rumberger. 2006. "Resource Needs for California's English Learners." Linguistic Minority Research Institute, University of California.

Hahn, J., P. Todd, and W. Van Der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design." *Econometrica* 69: 201-209.

Hill, L., J. Betts, B. Chavez, A. Zau, and K. Bachofer. 2014. *Pathways to Fluency: Identifying Success among Current and Former English Learners at the Elementary School Level*. San Francisco: Public Policy Institute of California.

Hill, L., M. Weston, and J. Hayes. 2014. *Reclassification of English Learner Students in California*. San Francisco: Public Policy Institute of California.

Hill, L, and I. Ugo. 2016. *High-Need Students and California's New Assessments*. Public Policy Institute of California.

Hill, L, J. Betts, M. Hopkins, M. Lavadenz, K. Bachofer, J. Hayes, A. Lee, M. Murillo, T. Vahdani, and A. Zau. 2019. A*cademic Progress of English Learners: The Role of School Language Environment and Course Placement in Grades 6-12*. Public Policy Institute of California.

Imbens, G.W., and T. Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics*142: 615-635.

Johnson, Angela. 2020. "The Impact of English Learner Reclassification on High School Reading and Academic Progress." *Educational Evaluation and Policy Analysis* 42 (1): 46–65. https://doi.org/10.3102/0162373719877197.

Kanno, Yasuko, and Sarah E. N. Kangas. 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 (5): 848-878.

Kim, J., and J. L. Herman. 2012. "Understanding Patterns and Precursors of ELL Success Subsequent to Reclassification." CRESST Report 818. National Center for Research on Evaluation, Standards, and Student Testing.

Linquanti, R. and H.G. Cook. 2015. Re-examining Reclassification: Guidance from a National Working Session on Policies and practices for Exiting Students for English Learners Status. Council of Chief State School Officers.

Los Angeles Unified School District. 2012. *English Learner Master Plan*.

Mavrogordato, M. and R.S. White. 2017. "Reclassification Variation: How Policy Implementation Guides the Process of Exiting Students from English Learner Status." *Educational Evaluation and Policy Analysis*, Vol 39, No. 2.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142(2): 698–714.

McLaughlin, M., L. Glaab, I. Hilliger Carrasco, June 2014. Implementing Common Core State Standards in California: A Report from the Field. Stanford University and PACE.

National Center for Education Statistics, 2019. *English Language Learners in Public Schools*.

Parrish, T., A. Merickel, M. Perez, R. Linquanti, M. Socia, A. Spain, C. Speroni, P. Esra, L. Brock, and D. Delancey. 2006. *Effects of the Implementation of Proposition 227 on the Education of English Learners, K-12, Findings from a Five-Year Evaluation*. Palo Alto: American Institutes for Research and WestEd.

Pope, N.G. 2016. "The Marginal Effect of K-12 English Language Development Programs: Evidence for Los Angeles Schools." Economics of Education Review, 53, 311-328.

Reyes, Marcela, and NaYoung Hwang. 2019. "Middle School Language Classification Effects on High School Achievement and Behavioral Outcomes." *Educational Policy*. https://doi.org/10.1177/0895904818823747.

Robinson, J. P. 2011. "Evaluating Criteria for EL Reclassification: A Causal-Effects Approach Using a Binding-Score Regression Discontinuity Design with Instrumental Variables." *Education Evaluation and Policy Analysis 33*(3), 267–92.

Robinson-Cimpian, J.P. and K.D. Thompson. 2016. The Effects of Changing Test-Based Policies for Reclassifying English Learners. *Journal of Policy Analysis and Management*, 35(2), 279-305.

San Diego Unified School District. 2009. *English Learner Master Plan*.

Saunders, W. M., and D. J. Marcelletti. 2013. "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 (2): 139-156.

Umansky, I.M., S. Reardon, K. Hakuta, K.D. Thompson, P. Estrada, K. Hayes, H. Maldonado, S. Tandberg, C. Goldenberg. October 2015. Improving the Opportunities and Outcomes of California's Students Learning English: Findings from School District-University Collaborative Partnerships. PACE.

U.S. Department of Education. 2014. *What Works Clearinghouse Procedures and Standards Handbook Version 3.0.* Washington, D.C.: What Works Clearinghouse, Institute for Education Sciences.

Warren, P. and P. Murphy. April 2014. "Implementing the Common Cores State Standards in California," Public Policy Institute of California.